

Problems with media gatekeepers

It is easy to point fingers at science writers and scientists when newspaper accounts of research throw patients and stock markets into unjustified frenzy. But editors, too, share responsibility.

To paraphrase Tolstoy, every overblown media furore causes unhappiness in its own way. Last week's excitements following a story about cancer research in the *New York Times* (see page 104) provide excellent examples of the good, the bad and the ugly in and around science journalism. Many are blaming the writer, Gina Kolata, for the problems her story generated. But whatever her responsibilities, there are other participants in the process whose roles need to be considered.

An important antecedent of this story was a paper published last November by Thomas Boehm and colleagues (also reported in the *New York Times*) describing the regression of tumours in mice following the administration of endostatin (*Nature* 390, 404; 1997). It also reported complete regression of Lewis lung carcinomas following joint treatment with endostatin and angiostatin — both are powerful agents that directly inhibit the formation of new blood vessels that feed tumour growth. An accompanying News and Views article celebrated the arrival of the two agents. Said author Robert Kerbel: "the results could herald a new era of cancer treatment. But that era could be years away...". There followed several reasons for that caution (*Nature* 390 335; 1997), including the fact that experiments in mice are by no means necessarily a harbinger of success in humans.

Last week's story contained essentially no new science, being apparently stimulated more by scientists' enthusiasm. The account as a whole was, on the surface, appropriately guarded. The word "cautious" (or its synonyms) appeared in the main headline, in the second paragraph, in a prominent picture caption about the team's leader, Judah Folkman, and several times elsewhere. The science and the uncertainties were explained in considerable detail.

But there were signs that the writer and editors wished both to have their cautionary cake and to eat it. The opening paragraph stated

that patients might be taking the new treatment within a year but failed to make it clear that this could be only on a last-hope, highly experimental basis. The story was full of the buzz of excitement in the laboratory and, crucially for its impact, overenthusiastic endorsements from James Watson and from Richard Klausner, head of the National Cancer Institute, both of whom subsequently repudiated or 'clarified' the words attributed to them. The positioning of the article emphasized the editorial double-think: the caveats considerably reduced the significance of the story, but its location high on the front page reinforced the allure of scientists' hopeful enthusiasm and belied that substantial uncertainty.

The consequences — soaring stock values of a biotechnology company involved in the research and thousands of telephone calls from desperate cancer patients pleading for non-existent treatment — were certainly regrettable and even tragic, but where does the blame lie? Typically (and the *New York Times* is no exception) a newspaper's front-page stories are picked at a conference of section editors who each advocate their stories. Does a tendency to hype arise in such selling? Are journalists tempted to oversell their story to news editors to gain such prominence?

The *New York Times* is probably the most scientifically aware and responsible of all US newspapers, but this tale demonstrates how even there the process can go awry. More generally, it is the job of news editors to ensure the appropriate balance of excitement and caution in stories they run. They, as well as science journalists, need to be able to apply acute sensitivity to potential misinterpretation by readers. And they can help to avoid bad consequences of their craft by a better understanding of the damage so often done by naive or, worse, disingenuous reporting of distinguished scientists' unbridled expressions of anticipation. □

Mirage of égalité persists

Conservative resistance to the reform of higher education is alive and kicking in France.

The thirtieth anniversary this month of the May 1968 student uprising on the streets of Paris is a powerful reminder to France's politicians of the sensitivities alive in universities. Claude Allègre, the minister for higher education, research and technology, has ambitious plans for reform. Progress comes with a report on the higher education system, commissioned by Allègre from an expert panel which, *inter alia*, proposes knocking the élitist and anachronistic *grandes écoles* off the pedestal where they have been for too long (see page 102). Its proposals for modernizing course structure in universities to reduce the huge student drop-out rate are also welcome.

Regrettably, the slim 46-page report fails to tackle seriously the central issue of how to give the universities greater independence from central government. One major obstacle is France's devotion to the mirage of an egalitarian university system, supposedly (but actually far from) uniform in quality, offering national degrees, and for-

bidden from selecting students on the grounds of ability.

The report is clearly only one element in the debate. Predictably, the notoriously conservative trade unions complain of a lack of consultation by the report's authors and the ministry. A more open and competitive university system would bring benefits to all, but the strong attachment of many French to a national system cannot be wished away by decree. The unions' ominous warning of "an *n*th reform imposed from above is doomed to failure" looks all too apt. In his year in office, Allègre has demonstrated an unenviable talent for rallying his opponents, and has often given the impression of abhorring genuine consultation. This may be a tactic intended to gain a strong starting position in negotiations. But if the government is to succeed in its admirable goal of modernizing French universities, it will at some point need to engage in a more subtle and well structured strategy for building a new consensus. □



UN agency chief seeks advice from environment groups

[BRATISLAVA] Only four months after taking over as head of the United Nations Environment Programme (UNEP), Klaus Töpfer, formerly Germany's environment minister, is already creating waves with his plans to reform the 25-year-old organization.

Töpfer said last week he wants UNEP to play a stronger role in the way countries implement the five UN conventions that relate to conservation, including taking a lead in developing relevant science advice.

Controversially, however, he wants much of this advice to come from non-government environment groups — including the International Union for the Conservation of Nature (IUCN), and the World Wide Fund for Nature (WWF) — as well as through the conventional route of other UN agencies.

In a speech to representatives of 172 countries at the fourth conference of the parties to the UN Biodiversity Convention in Bratislava, Slovakia, Töpfer said he was keen both to strengthen UNEP within the UN system and to bring non-government groups closer to the policy-making process.

Although few dispute the first goal, the initial reaction from delegates to the second was shock. A representative of a European Union state said bluntly that Töpfer's plans were "not even worth responding to". Another said more diplomatically: "We are in for some interesting times ahead."

But environment groups were delighted. Frank Vorhies, head of the IUCN economics unit in Geneva, said it was good news: "IUCN is well placed to play a role as UNEP's technical agency."

In a nine-page paper outlining his proposals, Töpfer says UNEP has a mandate from the UN general assembly to take a closer involvement in the conventions that UNEP helped to set up. These include the biodiversity convention, the Convention on



Hearing the news: last week's biodiversity meeting heard Töpfer (inset) spell out his proposals.

International Trade in Endangered Species (CITES), the convention on the conservation of migratory species of wild animals, and the wetlands convention.

At present, each convention has its separate agenda and reporting requirements, and its own advisory body of scientists. Scientists are drawn from different governments; they provide advice and help set research agendas. They report to a conference of country representatives for each convention, known as a Conference of the Parties.

Töpfer says he wants more "synergies" between conventions, by harmonizing scientific advice, programmes of work and reporting requirements. And he sees no reason why outside organizations with similar conservation aims should not be involved.

Science, says Töpfer, will be a key focus of his proposed reforms. Resources can be saved, he believes, if UNEP takes over from

individual secretariats the job of defining research methodologies and setting research priorities. UNEP, in turn, would seek help from non-governmental organizations (NGOs) as well as other UN agencies.

"I want to be as close as possible to organizations such as IUCN, and WWF, as well as other NGOs," he says. "The IUCN has an outstanding tradition. I intensely believe that they can be a part of the process."

Töpfer is in for a choppy ride. His plans to bring in environment groups to help implement the conventions will be strongly opposed, although it is unclear if he wants the groups to have a formal role in monitoring implementation.

Developing countries, in particular, are likely to consider this to be interference in their internal affairs. They are also likely to oppose plans to use environment groups for scientific advice.

The question of advice is particularly sensitive. At present, each convention takes scientific advice from a panel of scientists appointed by government delegations. The less developed countries prefer this arrangement, arguing that it gives them 'ownership' of the scientific advisory process.

Concern within the biodiversity convention that the advisory body was becoming too politicized (see *Nature* 391, 215; 1998) led some developed and advanced developing countries to suggest setting up independent panels of scientists in different disciplines, along the lines of the Intergovernmental Panel on Climate Change.

But even this suggestion is controversial, despite promises from its supporters that

Climate change centre signs up big firms

[WASHINGTON] Thirteen of the world's biggest corporations are to join a new information project, the Pew Center on Climate Change, which is intended to help raise public support for action to prevent global climate change caused by greenhouse gas emissions.

The centre will receive \$5 million a year from the Pew Charitable Trusts, although they will not receive any cash backing from the corporations, which include British Petroleum, Boeing, Toyota and Lockheed-Martin.

But the corporations — in a break from previous industry opposition to government action on climate change, which has been led by a group called the Global Climate Coalition — will put their names to advertisements that the Pew Center plans to publish this week.

The centre will be directed by Eileen Claussen, a former senior official of the US State Department and an important member of the US team that prepared to negotiate the Kyoto Protocol, signed last December.

Colin Macilwain

panel members would be drawn from as many countries as possible.

Some of the least developed countries question the need for scientific advice at an international level, given the biodiversity convention's emphasis on conservation in individual countries. They fear that new scientific panels will be dominated by scientists from developed countries, introducing a biased perspective to the advice they give.

Similarly, panels set up with environmental groups will be seen as partial to the environmentalist view. The role of IUCN may be particularly controversial, as many of its members appear to see conservation as more important than development.

"Conservation is a simple concept made

difficult by high-paid consultants," says Rabi Bista, special secretary in the ministry of forests and soil conservation in Nepal. "In my country, we know which areas need to be conserved. We have no difficulty at the professional level. Local people often know more than people like me in the cities. We don't need more committees [of scientists], we need local action."

Tewolde Berhan Egziabher, general manager of Ethiopia's environmental protection agency, believes a committee of outside scientists setting generic research priorities on issues that affect individual countries amounts to interference in sovereign matters. Individual countries, he says, should be left to commission their own research.

"Take research on the impact on biodiversity of deforestation," says Egziabher. "Whatever it finds, it will have implications for forestry policy. Science itself may be neutral. But there is no such thing as politics-free science at this level."

But Töpfer says he is confident that governments will take to his plans for UNEP and that its governing council of world environment ministers will endorse them when it meets at the end of this month.

He says he has the backing of the UN secretary-general Kofi Annan who, he says, appointed him to bring in new thinking. "Whether I want it or not, I need to deliver ideas. If they think I am wrong, they should show me the red card."

Ehsan Masood

Performance monitoring is 'last straw' for overworked NSF staff

[WASHINGTON] Complaints of overwork among scientific staff at the US National Science Foundation (NSF) may be reaching breaking point. The problem is exacerbated by new requirements for officials to monitor all NSF grants and projects.

A meeting of the physics and astronomy board of the National Research Council two weeks ago heard Bob Eisenstein, head of the NSF's mathematical and physical sciences directorate, predict a "trainwreck" at the science agency as officials struggle to manage projects and award investigator grants. "We don't have enough staff to do what we're trying to do," Eisenstein said.

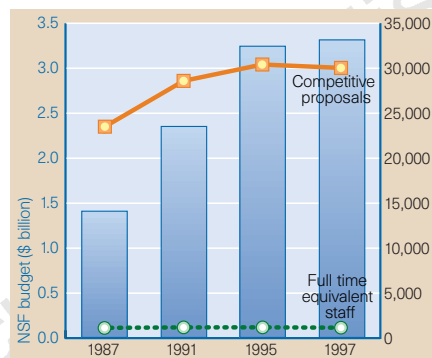
John Lightbody, head of the physics division in Eisenstein's directorate, added in discussion at the meeting that the Government Performance and Results Act (GPRA), approved by Congress in 1993, "will be how the trainwreck happens".

The act requires agencies such as the NSF to monitor the performance of its grantees more closely than ever before. It is imposing a large extra workload this year as officials implement detailed performance plans required by the legislation.

"People here are very worried about the workload implications of GPRA," Eisenstein said later. "It has the potential to consume enormous amounts of effort and energy." He added, however, that agency staff were still managing to process grant applications in "about the same time" as in previous years.

Eisenstein's comments reflect mounting apprehension at the NSF and other science funding agencies about the likely short-term consequences of GPRA at agencies where staff numbers have already been limited by the reluctance of either Congress or the White House to let them hire new staff.

In the case of the NSF, which funds most non-biomedical university research in the United States from its headquarters in Arlington, Virginia, the agency has been allowed to grow — but not to hire.



Left behind: NSF staff numbers have not grown in line with its budget and competitive proposals

Since 1987, the agency's annual budget has increased from \$1.4 billion to \$3.3 billion. The number of competitive proposals it receives has grown from 23,000 to more than 30,000, and the NSF is involved in ten times as many cooperative agreements as before. But it has the same number of full-time staff, 1,150, as ten years ago.

Indeed, the NSF often boasts that it spends just \$1 on administration for every \$20 it distributes in grants, less than most charities and foundations, and less than almost any other government agency.

Officials say that the staff's problems are compounded both by the increasing complexity of the proposals the NSF receives, and by the proliferation of special projects and 'cross-cutting' programmes. These address everything from the role of women in science to the special needs of regions that lack a strong science base.

"It's a real problem," concedes Richard Zare, professor of chemistry at Stanford University in California and outgoing chairman of the National Science Board, which oversees the NSF. "The ladders are close to breaking."

But Joe Bordogna, the acting deputy director of the NSF who is responsible for its GPRA effort, believes the agency can handle

the demands being placed on it. He says that the use of computer technology is helping.

Bordogna also points out that the agency has taken on more scientific staff on secondment from outside — 130 last year, compared with 70 in 1987 — and more contractors. In the early 1990s, the number of seconded staff grew sharply, but it has been held steady since 1994.

This year, the NSF is requesting an increase in administration funding of around 7 per cent, the first time in years that it has sought an increase above inflation for administrative costs.

The impact of administrative strains at the NSF on outside grant recipients is hard to discern. Bordogna says the agency is "probably getting more complaints" from scientists, but that this reflects a general state of flux in the research university system.

One ecologist who declines to be named complains that grant decisions normally taken in March have not yet been made this year, making it impossible for him to plan.

But others say there has been no change. "The NSF has always been horribly understaffed for the job it has to do," says Jim Brown, professor of biology at the University of New Mexico. "I don't know if the situation has become any worse."

Judy Sunley, an aide of Neal Lane, the NSF director, says the proportion of grant applications that the agency clears within six months "has been hovering around 50 per cent" but was just over 60 per cent last year. The NSF wants it to reach 70 per cent in 1999, and 95 per cent at an unspecified future date, Sunley says.

Staff on the congressional committees that oversee the NSF say understaffing has been a persistent and growing problem, but never the agency's top priority. But one former Republican staff member concedes that this might be because NSF knows that "it wouldn't be politically viable" to argue for more staff.

Colin Macilwain

Company aims to beat NIH human genome efforts

[WASHINGTON] A private initiative has been launched in the United States to use state-of-the-art technology to sequence the entire human genome within the next three years — four years earlier than the target of efforts being run by the federal government.

The Applied Biosciences Division of Perkin-Elmer, based in Norwich, Connecticut, is teaming up with J. Craig Venter, president of The Institute for Genomic Research



TIGR's aggressive approach looks set to pay off.

(TIGR) in Rockville, Maryland, to form a company devoted to sequencing rapidly more than 99.9 per cent of the genome's three billion base pairs. The ambitious goal is made possible, the principals say, by the launch of a \$300,000 gene sequencer by Perkin-Elmer, known as the 3700 DNA Analyzer. This combines increased sensitivity and automation to lower dramatically the cost and raise the rate of sequencing.

Venter says the cost will be reduced to \$0.10 per base pair (the government's effort costs about \$0.50). This would put the overall price of sequencing the human genome at between \$100 million and \$300 million — far below the \$3 billion government price.

The new company, to be located in Rockville, will house 200 to 230 of the new machines, which will be in operation by early 1999, say the executives. They say the operation will outstrip the capacity of all sequencing centres in the world combined, and that the company will act as a 'beta test' site for the new instrument.

Venter will leave TIGR to become president of the company, which is as yet unnamed. Its chairman will be Tony White, Perkin-Elmer's chairman, president and chief executive officer. Perkin-Elmer will own 80 per cent of the company, Venter and others the rest.

The company also intends to build an information-rich database adding information to raw sequence data which analyses it and thus makes it more valuable to scientists, companies and institutions. The goal, say the executives, is that the database should become "the definitive source of genomic and associated medical information".

It will, for example, include all-important data on common variations in DNA called polymorphisms. Ultimately, the principals say, the project will lay the groundwork for new diagnostics and therapies tailor-made to the needs of individuals.

Venter says that the company plans to make the database available to researchers for a reasonable fee, and to limit its patent claims to between 100 and 300 therapeutically important genes. The company will make the raw data themselves publicly available without charge. But the value, Venter says, will be in the database. "This is going to be an extremely high value database and we expect extremely wide subscriptions."

Some scientists have expressed fears that the company could at any time make the terms of access more onerous. Venter insists, however, that the data would not be "held hostage", saying that such a move would be "morally wrong".

The announcement was greeted cautiously but positively by the leaders of the government's Human Genome Project, who held a press conference with Venter and Michael Hunkapiller, the president of Perkin-Elmer's Applied Biosciences Division, on Monday (11 May) to respond to the news.

"We're hearing the discussion of a plan and not proof of the principle," said Harold Varmus, the director of the National Institutes of Health (NIH). He added that federal sequencing efforts at some dozen sequencing centres will continue "unabated for the moment". The NIH is the major federal player in sequencing the genome, with the Department of Energy playing a smaller role.

Francis Collins, the director of the National Human Genome Research Institute at NIH, called the venture "significant" but warned that its ability to meet its ambitious timetable remains to be demonstrated. It would be "vastly premature", he said, to alter current federal efforts in response to the news.

Ari Patrinos, the associate director of the Office of Biological and Environmental Research at the energy department, said the effort would "raise the bar" for government scientists, which would make them "even more productive".

The notion that a company could use new technology rapidly to outstrip the government effort — which has sequenced just 3 per cent of the genome so far — has put the government's role into a dramatically new light. It raises the question of whether Congress would continue to support the government project at the same level when a simultaneous effort was being carried out speedily with private money.

But Varmus warned that the proposed venture "will be one component factored into a very complex research agenda". Collins said that the plan to accelerate genome sequencing makes legislating against employment and insurance discrimination all the more urgent.

Meredith Wadman

Australia's research budget to remain flat for three years

[CANBERRA] The Australian government has failed to give higher priority to funding research, despite heightened lobbying by scientists and reforms of policy advisory mechanisms — including moving the science portfolio into the cabinet.

Budget details released this week reveal that public funding for the category of 'general research' will remain virtually flat over the next three years at A\$1,255 million (US\$797 million) in 1998–99, falling to \$1,247 million in 2001–02.

The third and final budget of the Coalition government delivers increases for medical research and marine science, and promises funding for a controversial new research reactor. But these have been offset by cuts elsewhere, in particular to the Australian Research Council, which provides support primarily for university research, and whose budget will fall from A\$445 million in 1998–99 to A\$390 million three years later.

Treasurer Peter Costello promised "wonderful opportunities" ahead as a result of the government's economic policies, which have moved the budget from deficit into surplus. But he gave no indication of any change of heart towards the funding of research as part of such prosperity, and his speech did not mention universities, science or technology.

Research gained one mention — an increase for medical research. Following intense lobbying by the biomedical research community, the National Health and Medical Research Council will receive A\$194 million in the coming year. Although its funding will fall to A\$179.5 million in 1999–2000, health minister Michael Wooldridge says this will be A\$50 million more than last year's forward estimates.

In his first year fully responsible for science, John Moore, Minister for Industry, Science and Tourism, secured some modest increases. The START programme of grants through the Industrial Research and Development Board spent A\$82 million in the financial year that is just finishing, A\$49 million less than its budgetary allocation. But a spokesman for Moore says the government's commitment to spending \$1 billion over four years on START will be "maintained".

The Australian Institute of Marine Science, based in Townsville, Queensland, will benefit from an extra A\$11.3 million over four years to replace one ageing research vessel, refurbish another and upgrade laboratory and other facilities. The institute's director, Russell Reichelt, greeted the funding as a "welcome boost to marine science".

But universities have received no relief for the cuts made to operational funds during the past two years.

Peter Pockley

France urged to open up élite education ...

[PARIS] France's élite *grandes écoles* should operate less in isolation from the rest of higher education and be integrated more into the university system through the creation of joint courses and degrees and the sharing of facilities such as libraries and laboratories.

This is the main recommendation of the report 'For a European Model of Higher Education' commissioned by Claude Allègre, the minister for national education, research and technology, from a 17-member panel that included Georges Charpak, who won the 1992 Nobel prize for physics, and the geneticist Axel Kahn.

The panel, chaired by Jacques Attali, who used to be adviser to the former president François Mitterrand and is himself a product of the *grandes écoles*, also included leading industrialists Pascal Brandys (chief executive officer of the biotechnology company Gen-set), Francis Mer (head of Usinor) and Jérôme Monod (head of Lyonnaise des Eaux-Suez).

The 200 or so *grandes écoles* — the best known of which are the Ecole Polytechnique, Ecole des Mines, Ecole des Ponts et Chaussées and Ecole Normale Supérieure — were set up in the eighteenth century primarily to supply top-flight engineers and scientists to the senior civil service. But their *raison d'être* is becoming increasingly obsolete, according to the report, given the

decreasing role of the state in industry and the economy.

Graduates of the *écoles* have served France well in the past within the state's massive technological programmes in such fields as nuclear energy, aerospace and high-speed trains. But economic needs have now shifted towards a higher education system that fosters contacts between scientists, entrepreneurs and small companies.

The report argues that greater integration of the *grandes écoles* into the university landscape is necessary to achieve this goal and to produce highly qualified graduates for the private sector.

One of the main weaknesses of the *grandes écoles* is in scientific research. Their courses do not provide for study and research towards doctorates, and the main centres of research are not in the *grandes écoles* but in the universities. Allègre consistently points out that this has resulted in France's senior civil servants often lacking a culture of research and science-based innovation.

To help remedy this situation, the report recommends the creation of a joint campus or 'poles' between neighbouring universities and *grandes écoles*, with students having access to courses in both — or being registered for joint degrees — while sharing libraries and other facilities.



Storm clouds gather: *grandes écoles* such as the Ecole Polytechnique may face changes.

The report recommends ending the monopoly enjoyed by the *grandes écoles* on recruitment to the state's scientific and technical corps, with posts in the senior civil service being opened up to university graduates for the first time. It also proposes that selection for entry to the *grandes écoles* should be diversified to allow a greater representation of social classes and to increase the proportion of students with a science education.

At present, the best school-leavers undergo a special year of training for the entry competitions to the *grandes écoles*, while the rest who have passed the *baccalauréat* go to university. The excellent facilities enjoyed by students at the *écoles* contrast sharply with the dilapidated and overcrowded conditions at many universities.

The report argues that the selection process has become too elitist, benefiting almost exclusively the children of top civil servants or wealthy industrialists, who have enjoyed a privileged education. This social inequality has widened in recent decades, according to the report, with most entrants to the *grandes écoles* now coming from just a handful of top schools.

To counterbalance this trend, the report recommends that students at universities and technical colleges, and not just school-leavers, should be allowed to apply to transfer to the *grandes écoles*. It also recommends that quotas should be fixed to guarantee an input of biologists and other science students.

The report's recommendations have been widely welcomed by university vice-chancellors and unions representing university students and staff. Etienne Gouyon, director of Ecole Normale Supérieure and himself a former university lecturer, also says he is "broadly favourable" to the report.

But he is concerned that opening up recruitment to the *grandes écoles* to universities might result in one-way traffic, depriving universities of their best students and impoverishing them further.

Declan Butler

... but university reform ideas meet resistance

[PARIS] Although the Attali report's recommendations for reform of the élite *grandes écoles* have been widely welcomed (see above), the fate of proposals for significant changes in the university system is uncertain. Past efforts ended ignominiously after protests by students and staff.

The report implicitly endorses the wish of the national education minister, Claude Allègre, to create competitive research-based universities. At present, universities are centrally organized, officially uniform in quality and offer national degrees.

The report recommends creating a national evaluation committee to rank individual university departments on research and teaching, and that more funding be awarded to the strongest. It also proposes increasing

institutions' autonomy by boosting the power of vice-chancellors and transferring buildings and land from the state.

At the same time, the report advocates making universities more accountable to economic and local needs through governing boards made up of local representatives and industrialists, to sit alongside the university board. In particular, governing boards could make nominations for vice-chancellor and participate in the election.

Both the conference of university vice-chancellors (CPU) and the unions are sceptical about this proposal. Unions argue that the current system, where the university president is elected by staff and students, is democratic and that outside interference would introduce economic interests at odds with

universities' role in teaching and fundamental research. Bernard Saint-Girons, head of the CPU, describes the proposal as "unsatisfactory".

More broadly, proposals to increase competition are likely to be extremely controversial. Many in France, and in particular unions representing university staff and students, remain strongly attached to a national university system with open access.

Claude Lécaillon, national secretary of SNESup, the main university teachers' union, describes the report's proposals as an "unacceptable" slippery slope to differential course fees and restricted access.

The education ministry last week declined to comment on the report's recommendations, although Allègre is reported to be pleased.

D.B.

German chemicals giant to focus on life sciences

[MUNICH] The pharmaceutical and chemical company Hoechst, one of the cornerstones of the German chemical industry since the Second World War, is to be restructured with a new focus on life sciences.

Plans to make the change were approved by the company's shareholders at their annual meeting in Frankfurt last week. But attempts to reduce the number of research scientists in a bid to streamline its research are meeting resistance from staff.

Hoechst will discontinue its traditional activities in industrial chemistry by the year 2000. At the same time, research at Hoechst Marion Roussel (HMR), the pharmaceutical company of Hoechst Holding, is to be restricted to core areas of drug design, contracting out other research activities, such as toxicology.

As a result, the number of scientific staff employed in research and development (R&D) in Frankfurt, HMR's largest research site, will be reduced by about 600 to 1,100. Similar reductions are planned at its research sites in Romainville, near Paris, and Bridgewater, New Jersey.

Hoechst has been under heavy financial strain in recent years, partly because of its broad spectrum of products and its low productivity in drug design. In the latter area, it has also been criticized for neglecting new methods, such as combinatorial chemistry, high-throughput screening of novel compounds and genomics.

In the past 12 months, Hoechst has been split up into eight separate companies, each with considerable independence within its field of business, but its difficulties have nevertheless continued.

"We cannot be at the top in everything we are doing," Jürgen Dormann, chairman of Hoechst's board of directors, told last week's meeting of shareholders. He said that regional markets for industrial chemistry products, such as standard synthetic materials, are increasingly being met by new Asian competitors and by oil companies.

In contrast to the cost-intensive chemical production of basic materials, biotechnology and genetic engineering would offer new knowledge-based opportunities in drug design, diagnostics, blood products, seeds, veterinary products and food additives, Dormann added.

Hoechst's pharmaceutical company, HMR, is by far the largest of the eight subsidiary companies. Although some describe it as the company's 'heart', it has also become its biggest headache. It spends DM2.4 billion (\$1.4 billion) a year on R&D, but its profits fell by almost 20 per cent in 1997.

"Productivity at HMR was much too low



over the past 20 years," said Dormann. He pointed out that, although an average of more than DM10 million a day had been spent on pharmaceutical research over that period, few marketable drugs had been developed.

Furthermore, only 5 per cent of HMR's turnover comes from drugs designed in the past five years, contrasting with an average of 20 per cent for its competitors. So HMR is to cut back on its research activities, confining them to promising projects in areas such as cardiovascular and metabolic disorders, rheumatology, immunology and cancer.

The company aims to reduce the time needed for the introduction of a new drug from 10–15 years to 6–9 years, and to quadruple the share of turnover from recently designed drugs within five years.

The generics part of the business will be sold, and research other than in the core areas of innovative drug design, such as toxicology, will be carried out in cooperation with external companies or research institutes, rather than by HMR's regular scientists.

The R&D department in Frankfurt will be designated the 'drug innovation and approval' arm, and the scientific staff will be reduced by a third by the end of 1999. HMR's global research budget will then be cut by DM460 million, DM95 million of which will be saved in Germany.

Felicitas Feick, a company spokeswoman, says the nature of HMR's research staff will be changed by taking on young researchers, especially biologists and geneticists.

But researchers at HMR have been opposing the dismissals. Some 7,000 employees protested against the cuts in January, and several hundred have continued to demonstrate in front of the Hoechst headquarters in Frankfurt each Monday.

Hoechst's managers are negotiating with the company's works committee and the unions. A final plan, including proposals for minimizing the staff reductions, will be presented next week.

Quirin Schiermeier

Internet research centre takes Intel inside China

[TOKYO] Intel, one of the world's largest manufacturers of semiconductors, plans to set up a US\$50 million research and development centre in Beijing to conduct research on the use of the Internet and related information technology.

The plan, announced last week, follows recent moves by several other leading multinational companies to set up research and development (R&D) centres in China's capital. Motorola, Hewlett Packard and IBM are said to be planning similar centres.

Intel plans to invest \$50 million during the next five years in the centre, which will carry out research with particular relevance to Chinese applications, such as speech recognition. It will also fund projects at China's leading universities and research institutes.

Despite the economic uncertainty in the region, several other multinational companies have also recently set up R&D centres in China. In April, for example, Procter and Gamble became the first multinational company to set up an R&D centre at the science park at Tsinghua University, one of China's leading universities, in Beijing.

The previous month, Bell Laboratories announced that it had set up a small R&D centre on Peking University's campus in Beijing, which will conduct joint research with the university.

"China has a deep pool of talented scientists and engineers, and there is a great opportunity to assemble a world-class team here to do some outstanding research," Andrew Grove, chairman and chief executive officer of Intel, said last week.

Intel's announcement has been welcomed by Chinese academics and researchers, who argue that the establishment of Intel's China Research Center will help to improve the level of fundamental research in high technology in China. "It represents a new trend for foreign investment in China," according to Peiyan Zeng, Chairman of China's State Development Planning Commission.

Some academics say that Intel must contribute actively to China's well-being, and not merely benefit from its resources. The centre should consider its relationship with the Chinese scientific community and provide practical help, argues Jiaxin Wang, deputy director of the State Key Laboratory of Intelligent Technology and Systems.

Research at the centre could stimulate new research at universities, he says. But the centre should promote collaboration with universities, and the company should use part of its profit to support education in China, says Wang, an expert on computer-aided pattern recognition, such as the recognition of Chinese characters. Richard Nathan

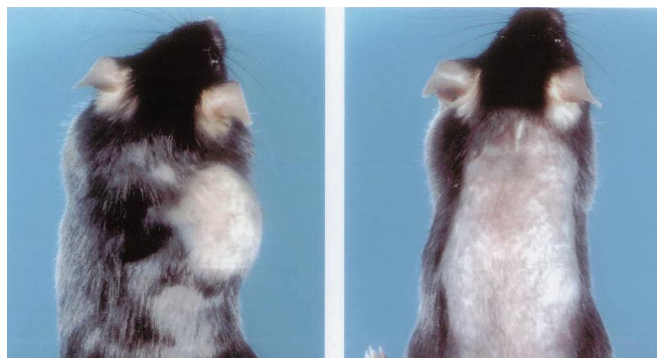
Cancer 'cure' article stirs up hot debate

[WASHINGTON] The presentation of a US newspaper story about innovative cancer research was last week criticized by cancer specialists, science-media watchers and biotechnology investors. They said the *New York Times* raised patients' hopes unrealistically and sent the stock market into a frenzy that was bad for both investors and industry.

"You have to think twice before you put a story above the fold on the front page about a drug and use the word cure when it really doesn't even exist in drug form today," says David Kessler, former commissioner of the Food and Drug Administration.

"People who should know better simply [got] involved in the hype. I would not exclude even members of the top echelons of the National Institutes of Health," says Spyros Andreopoulos, director emeritus of the office of communications at Stanford University Medical Center in Palo Alto, California, who has written widely on the interplay of science and the media.

The article, which ran on 3 May in the widely-read Sunday edition under the headline 'A Cautious Awe Greets Drugs That



Signs of hope: in Folkman's laboratory, the lung cancer tumour visible (left) was reduced after 12 days' endostatin therapy (right)

Eradicate Tumors in Mice', featured enthusiastic quotations from Nobel prizewinner James Watson and Richard Klausner, director of the National Cancer Institute. It described the work of Judah Folkman, a cancer researcher at the Children's Hospital in Boston, in whose laboratory the substances endostatin and angiostatin were discovered.

These natural substances work by choking the blood supply of cancers, and Folkman has said that in combination they eradicate all kinds of tumours in mice without inducing resistance or obvious side-effects.

Watson was quoted as saying that "Judah is going to cure cancer in two years". The article also quotes Klausner calling the drugs "the single most exciting thing on the horizon" for cancer treatment, and adding "I am putting nothing on higher priority than getting this into clinical trials".

Since the publication, both have modified their statements, or at least the story's presentation of them. The optimistic tone and prominent placement of the story — the top of the front page — sparked controversy.

Biotechnology investors pointed out that it ignored other important contexts, such as that a dozen or more laboratories are working on compounds that also seek to inhibit blood supply to tumours, and that many are further along in their research.

"There are 20 other companies with antiangiogenesis programmes. And most of them are more advanced really," says Stelios Papadopoulos, a biophysicist who is an investment banker specializing in biotechnology at PaineWebber in New York.

Cancer physicians complained they were flooded with calls from patients demanding the substances, which are not yet produced in sufficient quantities to launch human clinical trials and are thus years from availability, even if they surmount the hurdles.

Allen Lichter, a radiation oncologist at the University of Michigan who takes over as president of the American Society of Clinical Oncology at the annual meeting this coming weekend, says that the article was "cruel" in "inappropriately" raising hopes about a treatment that may never translate successfully from mice to men. "It builds cynicism that what people read about cancer research cannot be relied upon. That will accrue to our detriment over time," Lichter says.

But Nancy Nielsen, a spokeswoman for the *New York Times*, defends the handling of the story, arguing that the quotation from Watson made it "distinctive" and that the article described "an optimistic trend becoming apparent in scientists' thinking about a variety of recent lab developments".

At least one prominent scientist agreed. The molecular biologist and 1959 Nobel laureate Arthur Kornberg, of Stanford University,

Reporter backs away from lucrative book deal

[WASHINGTON] The day after the cancer research story appeared in the *New York Times*, the reporter responsible for it was negotiating with publishers about a book deal that an influential agent had promised would bring her a \$2 million advance.

News of the deal, first reported in the *Los Angeles Times* of 6 May, fuelled suspicion that the two events may have been jointly planned, although this is vigorously denied by both the reporter, Gina Kolata, and her agent, John Brockman.

Kolata quickly withdrew from the deal once issues about a possible conflict of interest had arisen. But the episode had already made its mark. "That someone might have written a story with the idea that this could lead to an enormously lucrative book deal is very unsettling," says Allen Lichter, president-elect of the American Society of Clinical Oncology.

Brockman says the deal

was his idea, and that he suggested it to Kolata the Sunday her story appeared. Kolata confirmed his account through a *New York Times* spokeswoman.

Brockman — who represents 175 scientists and science writers including Sir John Maddox, former editor of *Nature*, and Richard Dawkins — says that, on reading the story, he telephoned Kolata from his Connecticut farmhouse and told her: "I can get you two million dollars".

He says she at first resisted because of another book commitment, but by late Sunday had written the proposal and forwarded it to him by e-mail. Brockman sent the proposal to publishers at midnight, and by 9.30 on Monday morning had the first offer from a major New York publishing houses, he says.

On Tuesday, after several reporters had phoned Brockman enquiring about the book proposal, he says that he phoned Kolata, who

"was very upset", and told him: "I don't like this; I can see what's coming; I just don't think I can do this." She phoned Brockman later that day and asked him to withdraw the proposal.

Kolata referred calls to *New York Times* spokeswoman, Nancy Nielsen, who says that "after [Kolata] had some discussions with her editors, she decided to withdraw her book proposal", though the decision was entirely hers. The *New York Times* says it asks reporters not to write books on developing stories that they are still covering.

Brockman says that it would have been proper for Kolata to write a book on an issue that is newsworthy, because the "hypothetical" questions raised about conflict of interest disappear in the case of Kolata, who is "totally ethical".

Instead, he says, because of a misplaced outcry, "the ace reporter on biomedical issues is not going to write a book. Is that good?" **M.W.**

ty, says that the story deserved its front-page slot. "Something newsworthy has been found," Kornberg said, although he added he would have preferred the story to have been "a little more cautionary".

Watson and Klausner later toned down their remarks. Watson, in a letter to the newspaper published on 7 May, said "my recollection of the conversation [with the *New York Times* reporter] is quite different". But in the letter, he calls the work "the most exciting cancer research of my lifetime".

Wendy Goldstein, a spokeswoman for the Cold Spring Harbor Laboratory on Long Island, where Watson is president, said Watson was contesting the quotation attributed to him "because he feels a statement from him has offered what could very well prove to be false hope to a great many people".

The *New York Times* says it stands by the Watson quotation. But on 8 May it admitted it had imprecisely paraphrased Klausner by using the words "The National Cancer Institute has made the drugs their top priority" to describe his statement "I am putting nothing on higher priority".

In the *Los Angeles Times* of 6 May, Klausner — who could not be reached later last

week — sounded a cautionary note, saying "We have cured mice of cancer for decades — and it simply didn't work in humans."

Scientists argued last week that nothing in the *New York Times* story was news to those in the field. Folkman's early results with one of the substances, endostatin, in a particular mouse lung cancer were published in *Nature* last November (see 390, 404; 1997), and he has discussed results with both substances in scientific meetings for 18 months.

Nevertheless, the day after the news story appeared, shares in Entremed, the small company in Rockville, Maryland, founded to make the drugs, soared from \$12.06 to more than \$80 before closing at \$51.81. (They fell back to \$33.25 by the end of the week.)

Two other biotechnology companies Sugen and Magainin Pharmaceuticals also saw their share prices rise by 20 per cent and 38 per cent respectively. Both are working with similar compounds — known as angiogenesis inhibitors because they stop the growth of new blood vessels. But these compounds are already in human trials.

Biotechnology investors complain that the stock market's reaction damaged the industry. "It dilutes the impact of future

more definitive work by this company or others, and results in a potential misallocation of capital away from more proven treatments," says one New York investment banker, who calls the story "a bad thing for science and for biotechnology".

Experienced investors said the market's reaction — 23.5 million shares of Entremed were traded the day after the story was published — was driven by ordinary investors who engaged in momentum buying, in which people hurriedly buy a stock thinking that it may never be as cheap again.

Entremed's experience clearly stimulated some scientists into thinking about ventures of their own. David Nance, president and chief executive officer of Introgen Therapeutics, a company in Austin, Texas, that funds biotechnology start-ups, says that in conversations last week with researchers, all wanted to know: "If we go public fast, and publish this animal data showing this, is it worth a billion dollars these days?"

He told them "don't be ridiculous". Perhaps, he says, "if they were Nobel candidates and had reputations like Dr Folkman, maybe this would happen again in history. But don't count on it."

Meredith Wadman

EPA laboratories are 'still struggling' with change three years on

[WASHINGTON] Three years after the US Environmental Protection Agency (EPA) revamped its network of research laboratories, the labs are still struggling with the change, says the agency's Board of Scientific Counselors (BOSC).

Of particular concern is the scientific workforce, which it says is stretched thin and in some cases unsuited to its new tasks.

BOSC, set up in 1996 to advise EPA's Office of Research and Development (ORD) on its programmes and research agenda, relied mainly on written self-assessments by laboratory managers together with brief site visits by board members for its report.

Despite the speed with which BOSC worked — the review took only a few months — EPA science managers who have read drafts say they are impressed with its perceptive analysis. The final report goes to Congress and ORD this week.

BOSC looked at ORD's three national laboratories, each with dispersed divisions, and its two Washington-based 'national centres': the National Center for Environmental Research and Quality Assurance (NCERQA) and the National Center for Environmental Assessment. All are part of ORD, which conducts more than 80 per cent of EPA's research and has a budget of \$570 million this year.

Of the three main laboratories, the National Health and Environmental Effects Research Laboratory (NHEERL) has adapted best to ORD's new emphasis on risk-based

research, increased extramural grants and more rigorous peer review.

But reorganization has left the National Exposure Research Laboratory (NERL) short of scientific talent. In five years, it has lost half its budget, 28 per cent of employees and 45 per cent of contract researchers.

Now, says BOSC, the laboratory is "not self-sufficient in personnel and/or funding to carry out many of the large-scale research programs that fall under its mandate". Morale at the NHEERL improved after the reorganization; at NERL it "has not created any scientific excitement" and has not changed research methods.

Also facing manpower problems is the National Risk Management Research Laboratory (NRMRL) in Cincinnati, Ohio, which focuses on environmental cleanup and pollution prevention. The laboratory has skilled environmental engineers, says BOSC, but lacks many other professionals, from economists to microbiologists to behavioural scientists, needed for a broader role in risk assessment and management.

In recent congressional testimony, BOSC's chairman, Costel Denson, the University of Delaware's vice provost for research, said EPA needs "an orderly plan and program to enhance or redirect the skills of many of the scientists in ORD".

ORD managers recognize the need for new blood. The average age of its almost 2,000 employees is 48, and many are expected to retire in the next few years. EPA

asked the White House budget office for money to hire 200 postdoctoral researchers in 1999, but was granted enough for only 50.

Many EPA researchers feel hampered in collaborating with other scientists, the BOSC report says. They lack funds to attend meetings and believe — mistakenly, say ORD managers — that conflict-of-interest rules prevent them from interacting with outside scientists supported by EPA research grants.

Most of these grants are administered by the NCERQA under its Science to Achieve Results (STAR) initiative, which has grown in a few years from \$20 million to a \$100 million-a-year programme. This rapid growth, although considered positive, has taken its toll on NCERQA staff.

Grant administrators say workloads have increased so much that "they are now unable to dedicate the desired level of attention to the technical content of the research", the BOSC report says. NCERQA plans its own review of STAR this year.

But more than reviews, ORD needs money, says the agency's Research Strategies Advisory Committee. In an April report to the EPA administrator, Carol Browner, it pointed out that ORD's 1999 budget request was the lowest of the decade when adjusted for inflation. This comes as several major research projects are starting, including studies of particulate matter, endocrine disruptors, and microbial pathogens in drinking water.

Tony Reichhardt

Inserm restructuring increases political control, says union

[PARIS] A proposed reform of the French biomedical research agency Inserm was last week rejected by staff union representatives at a meeting with management. Although the most controversial elements of the decree — such as splitting Inserm into five departments — have been abandoned (see *Nature* 391, 110; 1998), unions are still protesting at what they claim is a bid to increase political control over the agency.

Controversy centres on the increased powers the agency's administrative council would have over strategic decisions on research directions and funding. These decisions were previously reached through agreement by the scientific community and Inserm management.

The unions maintain that assurances about the continued role of Inserm's scientific commissions in decision-making are insufficient, and that the proposed appointment of the administrative council chairman by an interministerial committee marks a shift towards greater central control.

Although Inserm is obliged to consult the unions, the ministry does not have to abide by the vote of last week's meeting. Inserm staff will have a last chance to lobby for changes to the decree at a meeting of the agency's administrative council next week.

ESA and CERN link on data and communication

[MUNICH] Two European research organizations have forged links in technical fields and communication. The European Space Agency (ESA) and the European Laboratory for Particle Physics (CERN) have agreed to learn from each other's experiences in areas such as the handling and networking of large amounts of data generated by large-scale experiments, and in communicating the results of their experiments to the public.

Following a meeting last week between Antonio Rodotà, director-general of ESA, and Christopher Llewellyn Smith and Luciano Maiani, outgoing and incoming directors-general of CERN respectively, the two organizations agreed to set up working groups to propose specific joint activities.

France and Germany reach biotech entente

[MUNICH] Germany and France have agreed to cooperate more closely in biotechnology research, particularly plant genome research. The agreement was reached by the two countries' respective research ministers, Jürgen Rüttgers and Claude Allègre, during a German-French summit last week.

A bi-national symposium will take place in September, to be attended by representatives of the two governments as well as industrialists and scientists from both countries, to discuss ways forward.

UK police chief calls for national DNA database

[LONDON] A senior member of the British police force has called for a national DNA database of the entire population, which, he says, would cut the time and costs of investigating crime. But government officials believe that such a database would be too expensive, and civil liberty groups see it as an infringement of privacy.

Peter Gammon, president of the Police Superintendents Association, said that a national database would enable criminals to be identified before they became serial offenders. He said the database could also be used in medicine, with police applying to the courts for permission to use it. "It could not be achieved overnight," Gammon told *The Independent* newspaper. "But I am asking [for] a cool and frank period of discussion."

'Decadal survey' to pinpoint star projects

[WASHINGTON] Joseph Taylor of Princeton University and Christopher McKee of the University of California at Berkeley will together chair the US National Research Council's next 'decadal survey' of priorities for astronomy and astrophysics. Taylor shared the 1993 Nobel prize in physics for his discovery of binary pulsars. McKee, a theorist, heads Berkeley's Space Sciences Laboratory.

The last decadal survey, released in 1991 and headed by Princeton's John Bahcall, has strongly influenced federal budget plans for both ground-based and space-based astronomy projects. The survey committee, which is still to be selected, will begin work this autumn, with a final report due by 2001.

Launch uncertainty ends plans for Neurolab 2

[WASHINGTON] The proposed Neurolab 2 space shuttle mission will not fly after all. The US National Aeronautics and Space Administration (NASA) had hoped to repeat the mission this autumn, giving space researchers from the United States, Europe and Japan more opportunity to study the nervous system's adaptation to weightlessness. Problems with research animals dying unexpectedly on the first flight (see *Nature* 393, 4; 1998) were seen as surmountable, and would only have required that some of the experiments be modified.

But uncertainty about launch dates for the first Russian and US components of the

international space station led NASA to scrap the idea of a Neurolab 2, and to reserve its autumn launch slot for a possible station assembly flight. After 15 years and 16 flights, the Spacelab programme has therefore come to an end.

Latin Americans form molecular biology body

[LONDON] A group of scientists from Latin America, Europe and the United States have set up an organization designed to develop molecular biology research in Spanish- and Portuguese-speaking countries. The Ibero-American Molecular Biology Organization (IMBO) was formally set up at a meeting in Chile last week.

The organization was initially proposed last year (see *Nature* 391, 524; 1998). The eventual aim is to set up an international molecular biology research laboratory in a Latin American country within five years. A more immediate priority is promoting exchanges between scientists from member countries, as well as a journal publishing primary research and articles examining issues affecting science and the ethics of science.

Spain's lack of career prospects laid bare



[MUNICH] Six young scientists, inspired by the British film *The Full Monty*, campaigned for better career prospects by stripping in a Spanish concert hall last week.

The scientists, members of the group PIC (Personal Investigador Contrado) which campaigns for career prospects for young researchers, performed in Madrid's Sala Galileo in front of more than a thousand.

In a sketch opening an evening of pro-science propaganda supported by several professional pop groups and dancers, the scientists acted out a scene where the only alternative to unemployment after years of training as postdocs was to become strippers.

In the finale they shed their lab coats to reveal on their backs one of PIC's demands, that the government should more than double research funding to 2 per cent of gross domestic product, the European average.

Italian defeatism unwarranted

Sir — Recent public declarations by the Nobel prizewinner Renato Dulbecco have cast a dark shadow over Italian science. In radio and television interviews he has lamented that the Human Genome Project is not getting funds, that the project is dead and that he may return to the United States.

Funding for the Italian share of the project has so far come from the “*progetto finalizzato*” of the National Research Council (CNR). Because of a major restructuring of the entire institution (whose budget is more than L1,000 billion (US\$560 million) a year) some ‘side projects’ have no doubt been neglected, including the Human Genome Project (which receives about L2 billion a year). Dulbecco was complaining about funding

for this project but the message has been perceived and amplified by the media as a major failure for all Italian science.

I am well aware of the existence of serious problems with science policy in Italy, as I unfortunately happen to be directly involved in one of them, but I also know that measures are being taken to improve the situation (see *Nature* **392**, 531; 1998).

Dulbecco’s advice (to leave the country) to me directly, and implicitly through his speech to the research community, is defeatist and will not help the younger generation of good, honest scientists who are trying to supplant a system still largely dominated by barons. Nor will it help agencies (such as Telethon and AIRC) that have finally introduced a serious system of

research funding, based on scientific peer review, and not on declarations through the media. Serious scientists should work for better management of science in Italy in general, not just worry when their own slice is threatened.

Those who know that the situation is different from that depicted by Dulbecco should make their voices heard and show that there is a way to change the system. Just leaving may be an option for some, but is certainly not the solution for Italian science.

Monica Zoppè

Istituto Tecnologie Biomediche
Avanzate CNR of Italy, LITA,
Via Fratelli Cervi 93,
20090 Segrate (MI), Italy
e-mail: zoppe@itba.mi.cnr.it

Funding needed for creative science

Sir — For science to contribute to breakthroughs to help people to live longer and healthier lives, it has to be creative. Pioneers are being squeezed out.

Too many scientists spend their careers repeating what their colleagues and/or competitors do in their field of research, with no new ideas, no new concepts, no creation. The race for grants and rapid publication, if possible in leading scientific journals with a high impact factor, probably account for this behaviour.

Awards like the Realizing Our Potential Awards (ROPA), recently extended in the United Kingdom (*Nature* **392**, 10; 1998), should be developed by governments elsewhere. By allowing the funding of ‘risky’ projects, they provide scientists with the opportunity to do creative research. Please, let us have more creativity, and more ROPA-like awards, for science in the future.

Bertrand Le Douarin

Université de Rennes 1,
Faculté de médecine CNRS UPR41,
2, avenue du Professeur Léon Bernard,
35043 Rennes Cedex, France
e-mail: bertrand@upr41.univ-rennes1.fr

New concepts of publication

Sir — I read with interest your leading article¹ about the withholding of data from “full and open access”², which is increasingly at issue across the sciences.

During the past four years or so, my colleagues and I have been developing

methods and procedures for the publication of research data in general and ecological data in particular. Our initial motivation came from the work of the Ecological Society of America’s FLED (Future-of-Longterm Ecological Data) Committee³ and more recently as part of a National Science Foundation grant supporting the development of a Web-based data management system for ecological analysis and synthesis⁴. This work continues to involve the society (publisher of *Ecology*) at the editorial, committee and research collaboration levels because of the unique and influential role professional societies and journals play in the debate about intellectual property rights in data.

There are recurring and fundamental issues limiting ‘full and open access’ to data that are intrinsic to institutionalized scientific research. Examples include the fear of being ‘scooped’ by someone using one’s data or inadequate attribution for one’s intellectual investment in a research programme resulting in the data, and the relationship this has to academic career advancement. The efforts of *Nature* and *Science* to address the issue of restricted access to data is a crucial and significant crack in the cultural mind-set that fosters, permits and even necessitates the withholding of research data.

The next steps along this path require fuller discussion and involvement of the scientific community along with the funding agencies and academic policy-makers. New ideas are needed. As digital library and data repository technology improves and the potential for broader and more rapid dissemination of data increases, consideration needs to be given to new concepts of publication. One approach

would be to raise data collections to the status of citeable entities in journals. Academic merit could then be obtained for the development and maintenance of data collections and it would be easier to recognize costs for data maintenance in grant proposals. Without such changes, there will continue to be little incentive for individual scientists to ensure the long-term quality and integrity of their often priceless data.

John Helly

San Diego Supercomputer Center,
University of California, San Diego,
La Jolla, California 92093, USA
e-mail: hellyj@sdsc.edu

1. *Nature* **391**, 617 (1998).
2. National Research Council. *Bits of Power: Issues in Global Access to Scientific Data* (National Academy Press, Washington DC, 1997).
3. Gross, K. et al. *Report of the Committee on the Future of Longterm Ecological Data* (1995) <http://esa.sdsc.edu/FLED/FLED.html>
4. Helly, J., Elvins, T., Sutton, D., Martinez, D. *Controlled Publication of Digital Scientific Data: An Example in Ecology*. In preparation.

Out of order

Sir — Blaxter *et al.* in “A molecular evolutionary framework for the phylum Nematoda” (*Nature* **392**, 71; 1998) incorrectly attribute to me “the view that vertebrate parasites evolved from arthropod parasitic ancestors”. The oxyurids of vertebrates have long been considered to have been derived from oxyurids in insects, but to extend this idea to the other orders of the nematode of vertebrates is weird indeed and cannot be attributed to me or to any expert on helminthology I can think of.

Roy C. Anderson

Department of Zoology,
University of Guelph,
Guelph, Ontario,
Canada N1G 2W1

The fraud of Abderhalden's enzymes

Earlier this century, the German biochemist Emil Abderhalden deceived the scientific world with his spurious 'defence enzymes'. Unless there is a change in clinical thinking, such a fraud could happen again.

Ute Deichmann and Benno Müller-Hill

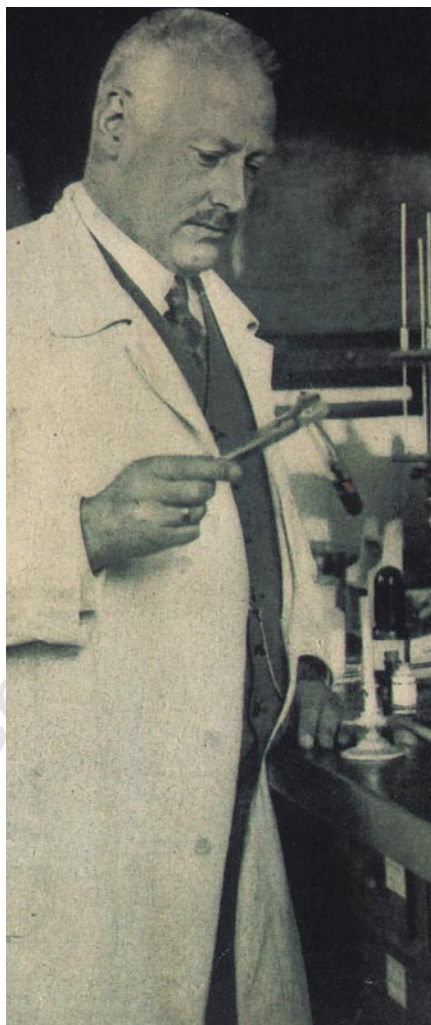
Is science a social construct, as some sociologists of science claim¹, or is its structure independent of social conditions? In the forefront of science, where everything seems possible, data may be misinterpreted and errors litter the path of science. In physics and chemistry such mistakes are usually quickly corrected by colleagues or the authors themselves. But what about fraud? Fraud assumes intention, which is difficult to prove. We think the deliberate invention and interpretation of data in science is a social construct. In physics and chemistry such social constructs are extremely rare; they are quickly detected and they have half-lives of just a few years. But we propose that this is different in medical science, in which science and social constructs may peacefully coexist.

We will demonstrate this by presenting the case of the non-existent *Abwehrfermente* (defence enzymes) created by Emil Abderhalden (1877–1950), who was professor of physiology and physiological chemistry at Halle University from 1911 to 1950, president of the Leopoldina—the oldest German academy of science—from 1931 to 1946, editor of several journals, and author of several books and more than 1,000 research papers².

Abderhalden's fraud

Abderhalden was born in 1877 in Switzerland. He received a medical education at the University of Basel, and in 1902 he went to Berlin to work with the great organic chemist Emil Fischer on the synthesis of peptides and the action of proteases, which are enzymes that break down proteins. In 1908 he became professor of physiology at the Tierärztliche Hochschule in Berlin, and three years later became professor of physiology and physiological chemistry at the University of Halle. He was due to become director of the Kaiser Wilhelm-Institut for physiology in 1914, but the First World War intervened. As a kind of compensation, the Kaiser Wilhelm-Gesellschaft financed his research with substantial grants until 1944.

Abderhalden's fame as a biochemist was achieved through two types of experiments. First, with Emil Fischer he began to synthesize and isolate peptides, and in his career he synthesized and isolated more than anyone else in Germany. Unfortunately, little use was made of them. In 1909 he published his first work on the second topic, the *Schutzfermente* (protection enzymes) or, as he called them later, the *Abwehrfermente*. In 1912 he published a book



Abderhalden: a convinced eugenicist.

about them, and considered them to be his most important discovery³; three new editions appeared before the end of 1914.

According to Abderhalden, animals and humans produce specific proteases, called *Abwehrfermente*, when they are challenged with foreign proteins. For example, the serum produced by pregnant women contains proteases specific to proteins of the placenta. The test for this claim is straightforward. Placenta is boiled, and the denatured, insoluble placental proteins are treated with serum from a pregnant woman. Peptides that arise through the action of the defence enzymes in the serum are dialysed and then identified by Biuret or ninhydrin reactions. Sera from non-pregnant women and men supposedly do not show this reaction.

This test intrigued gynaecologists and biochemists worldwide. Between 1912 and

1913, more than 25 papers from various gynaecological laboratories appeared that dealt with Abderhalden's pregnancy test, most of them with positive results³. In 1914, the directors of German university women's hospitals were asked by a medical journal to describe their experience with this test. Of the 15 that replied, all had more or less positive results, and none had negative results⁴.

The excitement increased. In the fourth edition of *Abwehrfermente* (1914), Abderhalden quotes 451 papers, many of them in non-German journals, which describe various uses of his test. As well as in pregnancies it was used successfully in three other contexts: the diagnosis of sarcomas and other carcinomas; the diagnosis of infectious diseases such as syphilis; and the diagnosis of psychiatric diseases such as schizophrenia. Cancer therapy using *Abwehrfermente* seemed just around the corner.

Why no one stopped him

We have to keep in mind that all these medical scientists deluded themselves: defence enzymes do not exist! It was a case of the emperor's new clothes: when everybody sees and admires his elegant clothes, just one child can destroy the social construct by pointing out that the emperor is naked. This 'child' was the German-Jewish biochemist Leonor Michaelis. In 1913 he had just published with Maud Menten the seminal paper on enzyme kinetics. Working in the biochemical laboratory of a hospital, he was asked by its director to establish the validity of Abderhalden's pregnancy test. He found that he and his collaborator were unable to repeat Abderhalden's experiments, despite spending a week in Abderhalden's laboratory in Halle. There was no difference between the sera of pregnant or non-pregnant women or between women and men: the pregnancy test did not work. In 1914, Michaelis and his collaborator published their negative results⁵; it marked the end of his academic career in Germany.

But Michaelis was not the only biochemist who could not repeat the results. Donald van Slyke from the Rockefeller Institute and Florence Hulton from the University of Pennsylvania failed too^{6,7}. In 1920 Jacques Loeb wrote⁸ to Michaelis, who was still in Germany: "Nobody speaks of the Abderhalden reaction any more in the United States and I am very much surprised to see that in his journal Abderhalden still continues that myth." The reply from Michaelis⁸ seems timeless: "In Germany one can succeed only when one presents practical,



applied science, however bad it may be. Anyone who wants to work on pure science is regarded a crank, and so he finally stops working". About Abderhalden he wrote⁸: "For me his type of work is disgusting. My position in Germany has suffered because of my opinion against his pregnancy test. There may be many who see through him, but nobody dares to say anything against him." Michaelis left Germany in 1922 to become visiting professor at a Japanese university, and later became a lecturer at Johns Hopkins University in Baltimore, Maryland, and then a member of the Rockefeller Institute in New York.

But how could Abderhalden continue with the *Abwehrfermente* from 1915 until his death in 1950? His strategy was simple and straightforward. He must have had collaborators who found what he wanted them to find (he dedicated the second edition of his book about the *Abwehrfermente* to his "faithful collaborators"). In general, he argued that the pregnancy test and other tests had worked in a large number of laboratories; so many scientists could not have deluded themselves. He conceded that in some laboratories the test did not work properly, and claimed that this demonstrated that the method was difficult, that it was not properly used, and — alas — that it was not perfect. So Abderhalden and his co-workers continued to streamline his method. The test became technically more and more complicated, and supposedly safer and safer. But the material to be tested became simpler. His research received a fresh impetus when he published a paper saying that the pregnancy test could work with urine, which was much easier to obtain than blood and supposedly contained more specific *Abwehrfermente* than blood. And so a second wave of users appeared.

How could Abderhalden continue with the *Abwehrfermente* from 1915 until his death in 1950? His strategy was simple and straightforward. He must have had collaborators who found what he wanted them to find

Sinister applications

In the 1930s and 1940s, a wide range of topics were analysed in half a dozen German institutes with the help of the *Abwehrfermente*⁹. There were tests for various forms of cancer; the final objective was cancer therapy. Tests for psychiatric diseases such as schizophrenia were developed; the *Abwehrfermente* were also used for effective shock treatment of psychotic patients. Tests were worked out to diagnose the various psychological types proposed by the psychiatrist E. Kretschmer¹⁰. (To test how patients dealt with fear, guns were fired behind their heads and pictures were taken; one schizophrenic patient is quoted as saying over and over again: "We want poison gas... why do they not give us poison gas?") The *Abwehrfermente* were used by Abderhalden's son Rudolf⁹ to diagnose various infectious diseases. They were even used to distinguish races of sheep⁹. Remember that none of these diagnostic tests could possibly work because the *Abwehrfermente* do not

exist, so the therapeutic value of the injections was dubious to say the least.

Some uses of the Abderhalden reaction were particularly disturbing. In November 1942, the human geneticist Otmar von Verschuer was appointed director of the Kaiser Wilhelm-Institute for Anthropology in Berlin. His former postdoc Josef Mengele joined him there in the winter of 1942–43. Mengele had been wounded in the war and spent some months convalescing in Berlin before moving in April 1943 to the Auschwitz concentration camp to become camp doctor. He must have discussed the scientific possibilities of Auschwitz with his teacher, because, when Mengele left, von Verschuer immediately applied for a grant from the Deutsche Forschungsgemeinschaft to finance Mengele's work in Auschwitz on the *Abwehrfermente* produced by members of various races deliberately infected with infectious diseases. They planned to use the Abderhalden reaction to demonstrate racial differences, and von Verschuer sent a technician to Abderhalden's laboratory in Halle to learn the technique¹¹.

Aware that the technique was not easy, von Verschuer called on Günter Hillmann to supervise the tests. Hillmann was an experienced biochemist who had worked in the laboratory of Karl Hinsberg, who had set out to improve the test of the *Abwehrfermente* for cancer cells; indeed, Hillmann himself had synthesized and tested a chemical that allowed a better quantification of the peptides released by the *Abwehrfermente*¹². But he had difficulties with Hinsberg, and in 1943 he moved to the laboratory of Adolf Butenandt, who had won the Nobel prize for chemistry in 1939.

Mengele sent blood samples from infected Jewish and Gypsy twins to Hillmann, who began analysing them. On 4 October 1944, von Verschuer wrote to a friend¹¹: "Precipitates have been prepared from the plasma of more than 200 individuals of various races, some twin pairs, some families. Abderhalden's method has been used and supplemented by a method newly discovered by Hillmann (who has joined us as collaborator). So very soon we can now begin our real research. The aim of our various efforts is now no longer to establish *that* the influence of heredity is important in various infectious diseases, but rather *how* hereditary factors act and what kind of events take place in their action."

A workshop on *Abwehrfermente* was held in 1947 in Tübingen, chaired by Butenandt. In a two-page report on the meeting¹³, Gerhard Mall, a collaborator of Kretschmer, wrote that Mall, Hinsberg, Kretschmer and Bersin had found specific defence enzymes. But Butenandt asked for the use of chemically homogenous proteins to re-test the claims.

That was not the last word on the *Abwehrfermente*, however. After Emil

Abderhalden's death in 1950, his son Rudolf declared that they were the perfect diagnostic tools to determine the optimal cell type for the *Frischzellen-Therapie* (fresh cell therapy) invented by Paul Niehans¹⁴, and so for a few years the Abderhalden reaction was used again. But two clinicians demonstrated that it made no difference whether the patient was healthy or sick; the sera reacted just the same¹⁵. We do not know when the *Abwehrfermente* finally disappeared as a diagnostic tool. Last year the *Frischzellen-Therapie* was outlawed in Germany by the Department of Health, but this ruling has been contested by doctors who claim it infringes the rights of both doctors and patients. The Supreme Court of Germany is pondering the problem, but this is likely to take a long time.

Behind the fraud

It is worth noting that Abderhalden was a convinced eugenicist. Given that his main scientific activity was based on self-deception and fraud, it is interesting that between 1922 and 1935 he edited a journal about ethics (*Ethik*). He wrote a textbook of biochemistry that appeared in 28 editions between 1906 and 1948 and was translated into four languages. He was president of the Leopoldina Academy from 1931 to 1950. After 1933, the notes "membership ended" or "membership extinguished" were secretly added to the file cards of the more than 90 Jewish members¹⁶ of the academy; the members were not informed, but they no longer received the academy's journal or invitations to events. All candidates for new international membership were discreetly vetted by the German foreign office to see whether or not they were Jewish¹⁶. After 1945, Abderhalden claimed that no Jewish member was ever expelled from the Leopoldina. Truth was not his business.

Historians of science have largely ignored Abderhalden and his *Abwehrfermente*. The biochemist Peter Karlson wrote¹⁷: "Emil Abderhalden certainly did not 'invent' the *Abwehrfermente*, he worked in many fields, he was a distinguished professor... and he certainly did not need to increase his fame through dubious publications... presumably one will have to classify the literature on *Abwehrfermente* as 'unconscious collection of wrong data', some kind of auto-suggestion." Theodor Wieland, a peptide chemist, calls Abderhalden¹⁸ "the founder of scientific biochemistry". About the *Abwehrfermente*, he wrote¹⁸: "Defence enzymes raised great hopes for theoretical biochemistry and also for practical medicine, which, however, in spite of intensive work, mostly with the participation of his son Rudolf, were not fulfilled. Thus they did not succeed in the isolation and characterisation of a defence proteinase."

The immunologist Otto Westphal told one of us (U. D., private conversation) that

his colleague Hans Brockmann wanted to work with Abderhalden on the *Abwehrfermente*: "Brockmann tested one or two systems, but was unable to reproduce them. He went to Abderhalden and told him that it worked the first time but not the second time. Abderhalden asked him why he repeated an experiment that worked well once. Brockmann left the institute immediately, considering Abderhalden to be a fraud." Westphal adds: "I had no doubt in the beginning myself, in fact I wrote a review on the *Abwehrfermente* in 1939. In 1942 or 1943 I spoke to Brockmann. The whole *Abwehrfermente* story was a fraud from beginning to end."

Never repeating an experiment that worked once, or discarding controls that did not work, is not science but pseudoscience or fraud. Abderhalden must have known this. It was his way of trying to seduce a young scientist to join him in the fake world of science as social construct. But Brockmann was a real scientist and fled. Westphal is listed as a participant at the Tübingen workshop on the Abderhalden reaction¹³. Was he critical? Did the participants use a double language that allowed the true scientist to abandon the non-existent defence enzymes and the believers to continue their social construct?

Can it happen again?

Abderhalden was a pure biochemist, but most of those using his method were medical, clinical biochemists. Such researchers often work during the day with patients in a hospital, and their experiments are confined to the afternoon or more likely the night. The senior author of research papers may be the director of a hospital. Clearly the universe of science is rather different from that of patients, who prefer to hear an optimistic diagnosis rather than the simple truth. It is difficult enough for the director of a large laboratory to validate all the experimental details of his collaborators, but for a physician, the director of a clinic who is responsible for many patients as well, this is clearly not possible. He must trust his collaborators, yet he is an authority. Truth may discreetly disappear.

In medical biochemistry, ideas or hope may be stronger than experimentally proven reality... The élite of today are students of the old élite, and learned the old values. Has medical, clinical science really changed that much?

Clinicians were offered the possible advantages of the *Abwehrfermente* for diagnostic breakthroughs. Most of them failed to admit that their tests did not work. Excellent non-clinical biochemists such as Butenandt and Kuhn kept silent too, at best stating that the specificity of the *Abwehrfermente* was not rigorously proven. The existence of the *Abwehrfermente* was seriously questioned only by Michaelis, van Slyke and Hulton, and the possibility of fraud was never mentioned in public. In medical biochemistry, ideas or hope may be stronger than experimentally proven reality. It is true that some antibodies have catalytic properties, but the question is whether the existence of specific defence proteases can be measured and proved in repeatable experiments. Here Abderhalden and almost all the people in the field failed.

At the time, Germany was regarded by many to be the leading country for medical science. The story is disturbing when we realise that it did not end in 1950 with the death of Abderhalden. The *Abwehrfermente* disappeared from the literature in the 1960s but nobody wrote a clarifying obituary. The élite of today are loyal students of the old élite, and they have learned and internalized the old values. Has medical, clinical science in Germany today really changed that much? We doubt it. The Brach-Herrmann-Mertelsmann affair¹⁹ provides a brief glimpse into the abyss of medical science in Germany. Will it be soon forgotten by the German medical élite, or will there be a real change in the spirit of true science? □

Ute Deichmann and Benno Müller-Hill are at the Institute of Genetics at Cologne University, Weyertal 121, 50931 Köln, Germany.

1. Latour, B. & Woolgar, S. *Laboratory Life. The Social Construction of Scientific Facts* (Sage, Beverly Hills, 1979).
2. Anonymous Z. *Vitamin. Horm. Fermentforsch.* 4, 1–18 (1951–52).
3. Abderhalden, E. *Die Schutzfermente* (Springer, Berlin, 1912).
4. Berichte über Krankheitsfälle und Behandlungsverfahren: Bedeutung der Abderhalden'schen Untersuchungsmethode. *Med. Klinik* 11, 453–455 (1914).
5. Michaelis, L. & Lagermark, L.v. *Deutsche Med. Wochenschr.* 7, 316–319 (1914).
6. Van Slyke, D. D., Vinograd-Villchur, M. & Losee, J. R. *J. Biol. Chem.* 23, 377–389 (1915).
7. Hulton, F. s. *J. Biol. Chem.* 25, 163–171 (1916).
8. Library of Congress (1920, 1921).
9. Harrer, G. Z. *Vitamin. Horm. Fermentforsch.* 1, 484–503 (1947–48).
10. Kretschmer, E. & Mall, G. *Fermentchemische Studien zur klinischen und konstitutionellen Korrelationsforschung speziell zur psychiatrischen Endokrinologie* (de Gruyter, Berlin, 1941).
11. Müller-Hill, B. *Murderous Science* (Oxford University Press, 1988).
12. Hillmann, G. *Hoppe-Seyler's Z. Physiol. Chem.* 277, 222–232 (1943).
13. Mall, G. Z. *Vitamin. Horm. Fermentforsch.* 2, 47–48 (1948–49).
14. Abderhalden, R. in *Die Zellulärtherapie* (ed. Niehans, P.) (Urban & Schwarzenberg, Munich, 1954).
15. Kanzow, U. & Schulten, H. *Die Medizinische* 13, 447–450 (1957).
16. Gerstengarbe, S. *Leopoldina (R3)* 39, 363–410 (1994).
17. Karlson, P. *Naturwissenschaft. Rundsch.* 39, 380–389 (1986).
18. Wieland, T. & Bodansky, *The World of Peptides. A Brief History of Peptide Chemistry* (Springer, Berlin, 1991).
19. Schiermeier, Q. *Nature* 389, 105 (1997).

Acknowledgements. This work was supported by Deutsche Forschungsgemeinschaft.

Real and realistic quantum computers

David P. DiVincenzo

The first real quantum-mechanical computations have now been made. But current techniques cannot be scaled up to produce useful quantum computers. A radical scheme, using semiconductor physics to manipulate nuclear spins, could be the answer.

In a quantum computer, calculations are made by the controlled time evolution of a set of coupled two-level quantum systems (qubits). Since Shor¹ showed, four years ago, that some hard computational problems — such as factoring integers — become tractable in a quantum computer, the desirability of producing a real one has hardly been a matter for serious debate. It is a daunting task, but physicists and chemists have begun to confront the challenge of building the necessary hardware — and two papers in this issue show that we are making progress.

On page 143, Chuang *et al.*² report one of the first realizations of a simple but uniquely quantum-mechanical algorithm (see box, overleaf), achieved by nuclear magnetic resonance (NMR) spectroscopy on the small organic molecule chloroform. This achievement does not indicate a clear route to large-scale quantum computation — but on page 133, Kane³ lays out an audacious vision for how that might be achieved by a radical new design, combining NMR techniques with the semiconductor physics that underpins present-day computer technology.

In NMR spectroscopy, each qubit is realized in the spin orientation of an individual atomic nucleus — the direction of the nuclear magnetic dipole. Each dipole can either reinforce or oppose an externally applied magnetic field. The first state has lower energy than the second, and the state can be changed by the absorption or emission of a radio-wave photon of the right energy. This coupling to the radiation field also provides a way of externally measuring the spin states of a large ensemble of identical nuclei on different molecules. Different nuclear spins on the same molecule interact through the magnetic dipole force, permitting basic logic functions involving these qubits to be carried out.

Despite its impressive success, the bulk NMR approach to quantum computing will be hard to scale up beyond the quantum-parlour-game level. First, the simple approach of distinguishing qubits on a molecule by their chemical identities becomes impossible for larger molecules. Second, the technique requires about 10^{18} chloroform

molecules in solution to measure successfully the result of the computation; such massive redundancy becomes unfeasible if the computation itself requires thousands or millions of bits. Lastly, the simplicity of employing a simple organic solution at ambient temperature presents a basic problem for scaling up. Successful computation requires a properly initialized starting state (with all spins down, for example); but in thermal equilibrium at room temperature, the probability that all the spins start out down in an N -spin molecule decreases exponentially with N .

By marrying the NMR approach with solid-state physics, Kane provides a vision of how to solve these problems and make large-scale quantum computation a real possibility. The heart of his proposal is the repetitive device structure shown in Fig. 1. It is made of exactly the same component parts as present-day computer devices — a semiconductor crystal (silicon), dopants (phosphorus), an insulating layer and metal gates. However, its operation bears no resemblance to that of a transistor. For example, the answer to the standard question, “where does the current flow when the device switches?” is, “it doesn’t”. Kane’s computer is the ultimate no-moving-parts machine.

Not even the electrons move, at least not much.

The qubit in the Kane scheme is again a nuclear spin, that of a ^{31}P dopant buried inside the silicon crystal directly beneath a metal contact (an ‘A gate’). Phosphorus in a silicon host is an electron donor, meaning that it has one outer electron which, at room temperature, moves freely in the crystal lattice, carrying the current for transistor functions. But at a temperature of 100 mK, this electron is weakly bound to the phosphorus ion, and its spin can influence the state of the nuclear-spin qubit. So by applying a voltage on the A gate, the bound electron can be polarized, changing its overlap with the donor nucleus. Now, instead of the external magnetic field, it is the interaction between electron and nucleus that determines the relative energies of the two nuclear spin states, and therefore what radio-wave frequency is required to flip the spin. So a radio pulse can be used to selectively change the state of only the nucleus whose electron has been polarized. Other gates (‘J gates’) influence the overlap of neighbouring electrons, mediating an indirect coupling between ^{31}P qubits, allowing any operations needed to carry out a quantum algorithm⁴. This approach eliminates the first problem of present-day NMR quantum computing, by making particular spins addressable not by their chemical identity, but by their location under a long linear array of metal gates.

Kane proposes an elegant scheme for eliminating the second problem, by reading out the state of the *individual* nuclear spins. Instead of a direct external measurement of the spin state, which has proved very difficult⁵, Kane describes a chain of interactions connecting the nucleus’s spin to the macroscopic world. First comes an operation that maps one of the ^{31}P spins

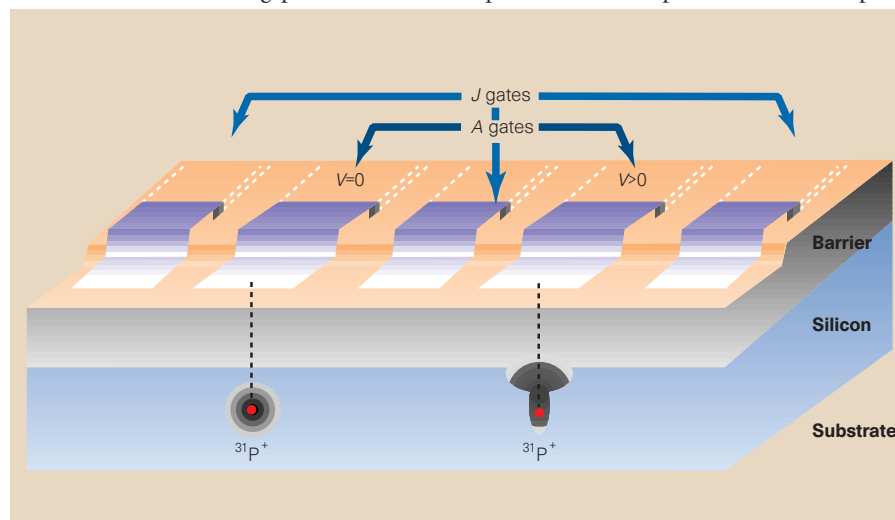


Figure 1 The device structure in Kane's proposal³ for a silicon-based quantum computer. The isolated dopant ^{31}P nuclear spins are the qubits. The overlying metal gates implement quantum logic operations by affecting the shape and overlap of the electron wavefunctions surrounding each phosphorus nucleus. More qubits are added simply by repeating this basic structure.

Seeing both sides

The algorithm implemented by Chuang *et al.*² was first considered by David Deutsch⁷ in his pioneering work on quantum algorithms in the mid-1980s. It provides a quantum solution to a simple parlour game: I hand you a coin – is it fair or fake? That is, does it have a head and a tail, or are both faces the same? You can certainly tell with two glances, first at the front

of the coin and then at the back, but can you tell with only one? Obviously not – unless your glances can be quantum-mechanical. In the experiment, the qubit indicating which glance is to be made (0 for front, 1 for back) is embodied by the spin-up/spin-down levels of the hydrogen nucleus in each molecule of chloroform (HCCl_3), whereas the outcome of the glance (heads/tails) is held by the states of

the ^{13}C nucleus. The constitution of the coin is encoded in a simple initial NMR pulse sequence. The essence of the algorithm is to place the 'glance' qubit in a quantum superposition of the 0 and 1 (front and back) states. As the experiment shows, an appropriately designed NMR measurement can then determine 'fair or fake' in a single shot.

D.P.D.

to the spin of an electron pair. Usually, all the electron spins are aligned with the external magnetic field, but when the overlap of two neighbouring electron orbitals is made very strong by a large swing of the J -gate voltage, it becomes more energetically favourable for that pair of electron spins to become anti-aligned. Whether the electrons go into this state during the voltage sweep, however, or remain in the metastable high-energy state in which their spins are aligned, is determined by the direction of whichever ^{31}P spin is coupled most strongly to its electron (as determined by the A -gate voltage). It is then possible to measure the spin state of this electron pair externally: if their spins are aligned, the Pauli exclusion principle forbids both electrons from hopping onto the same phosphorus, whereas that process is permitted if they are anti-aligned. These possibilities may be distinguished by measuring the capacitance between the neighbouring A electrodes.

This measurement protocol can solve the third problem of present-day NMR quantum computing by initializing the nuclear spins in an almost perfect spin-down state. The procedure is simple: as nuclear spins can be measured individually, any 'up' states can be flipped to 'down' with an NMR pulse. Thermal spin fluctuations are already reduced by operating at 100 mK rather than 300 K, but this trick further reduces the effective temperature — and by an enormous factor.

Kane's proposal satisfies all the acknowledged theoretical criteria for the construction of a quantum computer⁴. Still, he recognizes that the fundamental and engineering obstacles to implementing the scheme are vast. No existing materials-preparation technology will place an array of individual phosphorus atoms at desired spots in the interior of a perfect crystal; nor are there

yet systems free from defects in the semiconductor and the overlying oxide layer. But these are problems whose solution will be more and more necessary for the advancement of conventional computer technology, so great resources will be devoted to tackling them.

Not only better technology, but better understanding of fundamental quantum processes, will be needed to make this computer go. The quantum states of ^{31}P lose their coherence only very slowly in the bulk crystal, but whether the decoherence rate

will be small enough⁶ in the device geometry of Fig. 1 is almost impossible to predict. Decoherence from interactions with other spins could mean that isotopic purification of the silicon crystal may be necessary, as silicon naturally contains a few per cent of the spin-containing isotope ^{29}Si . Finally, the motion of other defects in the crystal may produce random voltage offsets that must be measured and compensated for by complex external circuitry.

The drive to construct ever more powerful computing devices has challenged some of the most inventive technological and scientific talents of the past two generations. It appears that, although very deep, this well is far from dry. Proposals like Kane's show that radical new departures in the way we do computing may still be possible, despite 50 years of non-stop revolution. □

David P. DiVincenzo is at the IBM T. J. Watson Research Center, PO Box 218, Yorktown Heights, New York 10598, USA.

e-mail: divince@watson.ibm.com

1. Shor, P. in *Proc. 35th Annu. Symp. Foundations of Computer Science* (ed. Goldwasser, S.) 124–134 (IEEE Computer Soc., Los Alamitos, CA, 1994); *SIAM J. Comput.* **26**, 1484–1509 (1997).
2. Chuang, I. L., Vandersypen, L. M. K., Zhou, X., Leung, D. W. & Lloyd, S. *Nature* **393**, 143–146 (1998).
3. Kane, B. E. *Nature* **393**, 133–137 (1998).
4. DiVincenzo, D. P. & Loss, D. *Superlattices and Microstructures* **23**, 419–432 (1998).
5. Rugar, D., Yannoni, C. S. & Sidles, J. A. *Nature* **360**, 563–566 (1992).
6. Preskill, J. *Nature* **391**, 631–632 (1998).
7. Deutsch, D. *Proc. R. Soc. Lond. A* **400**, 97–117 (1985).

Gene expression

Transcription sans TBP

Lynne M. Apone and Michael R. Green

In the field of transcription there is little room for dogmatism — principles that were once deemed sacrosanct can be violated with impunity. The latest example of this is illustrated on page 187 of this issue, where Wicczorek *et al.*¹ show that a factor that was thought to be essential for transcription, the TATA-binding protein (TBP), is, at least under some circumstances, dispensable.

RNA polymerase II — the enzyme that transcribes protein-coding genes — cannot initiate transcription accurately without the assistance of a large group of auxiliary general transcription factors. These include TFIIA, TFIIB, TFIID, TFIIIE, TFIIF and TFIIH², of which TFIID has always commanded special attention. TFIID comprises TBP and a set of highly conserved 'TBP-associated factors' (TAF_{II}s)³. The TBP component recognizes and binds to the TATA box (the signature sequence of the promoter) to initiate assembly of the transcription complex. Other general transcription factors such as TFIIB bind to this TBP platform

and, through a series of protein–protein interactions, RNA polymerase II is recruited to the promoter.

The TAF_{II} components of TFIID were thought to have a different, but equally important, activity — enabling promoter-specific activator proteins to stimulate transcription³. So, TFIID seemed to provide three indispensable functions: promoter recognition through the TBP–TATA-box interaction; complex assembly through interactions between TBP and the general transcription factors; and activator-directed transcription through the TAF_{II}s.

The first blow to this traditional view of TFIID came several years ago. Studies in yeast revealed that a variety of genes were transcribed normally *in vivo* in the absence of functional TAF_{II}s^{4–6}. This result, which seemed heretical at the time, indicated that TAF_{II}s were not obligatory for the general function of promoter-specific activators.

Wicczorek *et al.* now present a different assault on the traditional picture by showing the dispensability of the TBP subunit, the

cynosure of TFIID. They used an antibody to a specific TAF_{II} — TAF_{II}30 — to purify complexes from human cells. Unexpectedly, one of these TAF_{II}30-containing complexes could not be immunoprecipitated with an anti-TBP antibody. The authors found that it contained TAF_{II}30 and a subset of TAF_{II}s, but no TBP. Accordingly, they named it TBP-free TAF_{II}-containing complex (TFTC). In both relatively crude and highly purified systems, TFTC could support transcription of a promoter containing a conventional TATA box, as well as a so-called TATA-less promoter. Activity of the TATA-less promoter is mainly directed by an 'initiator element' at the transcription start site (Fig. 1). The TFTC-directed transcription was not inhibited by an anti-TBP antibody, and TFTC could also support the response to a transcriptional activator.

Although the studies of Wicczorek *et al.* are certainly unexpected, hints of their main conclusions can be found in several previous reports. In particular, transcription from a promoter containing a binding site for the activator YY1 requires only three components⁷: YY1, TFIIB and RNA polymerase II. The discovery of a distinct, functional, neural-specific TBP indicates that the standard TBP is unnecessary, at least in some circumstances⁸.

How can we reconcile the dispensability of TBP with the idea that it is essential in promoter recognition and complex assembly? Binding of TFIID to the promoter involves a TBP–TATA–box interaction. But many contacts are also made between the TAF_{II}s and the core promoter⁹ (the site at which the general transcription factors assemble). Indeed, in DNA-footprinting experiments, TFTC binds to core promoter sequences surrounding the TATA box, but not the TATA box itself¹. This indicates that TAF_{II}s are sufficient for recognition of at least certain core promoters. Such interactions between TAF_{II}s and the core promoter may also be the basis by which TAF_{II}s can direct selectivity for the core promoter *in vivo*¹⁰. The dispensability of TBP for the assembly of transcription complexes may be explained by direct interactions between some TAF_{II}s and certain general transcription factors³.

The finding that transcription can occur without interactions between TBP and a promoter is reminiscent of the situation with eukaryotic RNA polymerases I and III, and RNA polymerase II-directed transcription of TATA-less promoters and small nuclear RNA genes. In these systems, auxiliary transcription factors are involved, one of which contains TBP¹¹. But the activity provided by TBP is unclear — because the promoters lack a canonical TATA box (Fig. 1), TBP is unlikely to act through a sequence-specific promoter contact. Rather than providing an activity that is critical for the initiation of transcription, perhaps TBP

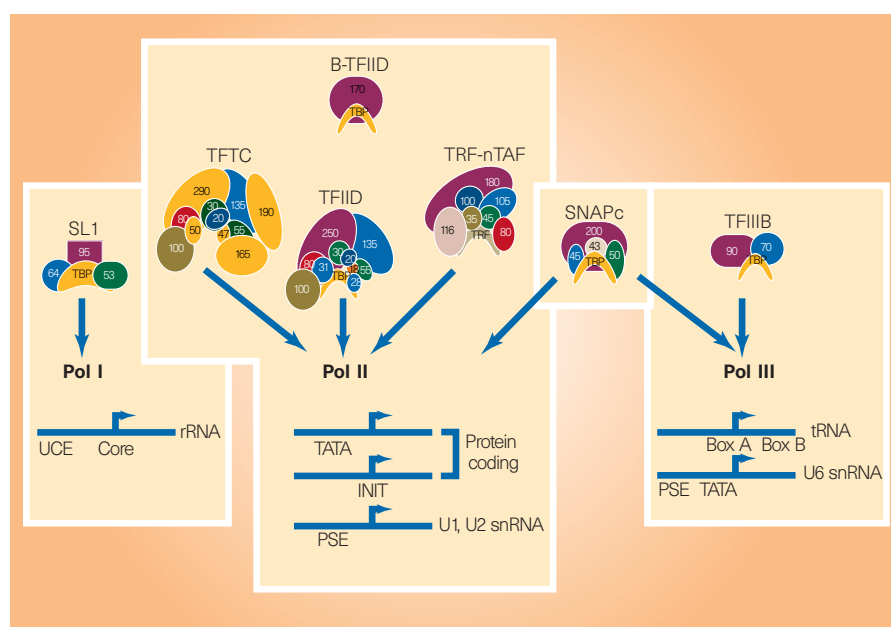


Figure 1 Complexes that contain TATA-binding protein (TBP) and TBP-associated factors (TAF_{II}s) in transcription by the three eukaryotic RNA polymerases. From left to right, transcription of ribosomal RNA (rRNA) genes by RNA polymerase I (Pol I) requires SL1. Transcription of protein-coding genes by RNA polymerase II (Pol II) requires TFIID, TRF–nTAF and TBP-free TAF_{II}-containing complex (TFTC), which has just been identified by Wicczorek *et al.*¹. B-TFIID comprises TBP and Mot1, and is presumed to regulate the availability of TBP. SNAPc is involved in transcription of small nuclear RNAs (snRNAs) by RNA polymerases II and III. Transcription of transfer RNA (tRNA) genes by RNA polymerase III (Pol III) requires TFIIB. TAF_{II}s are labelled according to their apparent molecular weights.

acts mainly during assembly of the multi-subunit complex. The presence of a common factor (TBP) in each of these complexes may coordinate the different classes of transcription.

Wicczorek *et al.*¹ have clearly shown that TBP is dispensable, but what of the TATA box? In a typical promoter, mutation of the TATA box generally abolishes transcription. So, if TFTC is a bona fide complex, we would expect it to be present in crude transcription extracts and to support the activity of a mutated promoter. These considerations raise the possibility that TFTC is not a natural complex, but that it was generated during purification. Alternatively, another component may contact the TATA box during TFTC-directed transcription. Crystallographic studies have shown that binding of TBP to the TATA box dramatically distorts the core promoter DNA³. An implication of the new results is that the structure of DNA may not need to be altered for initiation of transcription.

The results of Wicczorek *et al.*, taken in conjunction with those from other studies (for example, ref. 8), indicate that many TBP- and TAF_{II}-containing complexes are involved in transcription directed by RNA polymerase II. Each of these complexes represents a precise combination of common and unique components, so deciphering the pathways and regulation of complex assembly should prove fascinating. Various

questions remain. Why does a single TFIID not suffice? What promoters do these various complexes act on? And how do the various TFIID complexes recognize their target promoters? The discovery of TFTC reinforces a well-established theme in the field — the existence of numerous multi-subunit general transcription factors. Such complexity seems curious if, as has been suggested, many of these multi-subunit complexes are pre-assembled into a single holoenzyme. □

Lynne M. Apone and Michael R. Green are in the Howard Hughes Medical Institute and Program in Molecular Medicine, University of Massachusetts Medical Center, 373 Plantation Street, Worcester, Massachusetts 01605, USA.

e-mail: michael.green@ummed.edu

1. Wicczorek, E., Brand, M., Jacq, X. & Tora, L. *Nature* **393**, 187–191 (1998).
2. Zawel, L. & Reinberg, D. *Annu. Rev. Biochem.* **64**, 533–561 (1995).
3. Burley, S. K. & Roeder, R. G. *Annu. Rev. Biochem.* **65**, 769–799 (1996).
4. Apone, L. M., Virbasius, C. A., Reese, J. C. & Green, M. R. *Genes Dev.* **10**, 2368–2380 (1996).
5. Moqtaderi, Z., Bai, Y., Poon, D., Weil, P. A. & Struhl, K. *Nature* **383**, 188–191 (1996).
6. Walker, S. S., Reese, J. C., Apone, L. M. & Green, M. R. *Nature* **383**, 185–188 (1996).
7. Usheva, A. & Shenk, T. *Cell* **76**, 1115–1121 (1994).
8. Hansen, S. K., Takada, S., Jacobson, R. H., Lis, J. T. & Tjian, R. *Cell* **91**, 71–83 (1997).
9. Oelgeschlager, T., Chiang, C.-M. & Roeder, R. G. *Nature* **382**, 735–738 (1996).
10. Shen, W.-C. & Green, M. R. *Cell* **90**, 615–624 (1997).
11. Geiduschek, E. P. & Kassavetis, G. A. *Curr. Opin. Cell Biol.* **7**, 344–351 (1995).



100 YEARS AGO

An automatic telephone exchange system, which does away with the necessity for the staff of skilled operators at present exquired at exchanges, is being introduced into this country from the United States by the Direct Telephone Exchange Syndicate. Instead of ringing up the central station, requesting the attendant to put him in communication with the person to whom he wishes to speak, and waiting while the required alterations are made on the switch-board, the subscriber to an exchange worked on the automatic plan is himself able to connect his telephone with that of any other subscriber without the intervention of a third person. Each subscriber has upon the front of his instrument a circular disc pivoted at the centre, and having one-half of its circumference inscribed with figures from 0 to 9. If he wishes to communicate with another, he sets the disc so that the number of the other subscriber appears upon the dial, and he then finds his telephone in circuit with that of the person whose number he has indicated by his disc. When he has finished his conversation he simply hangs his receiver on its hook. Immediately, the switch which represents him at the exchange returns to its normal position, and communication is cut off.

From *Nature* 12 May 1898.

50 YEARS AGO

It is known that the spherical aberration of electron lenses sets a limit to the resolving power of electron microscopes at about 5 Å. ... The new microscopic principle described below offers a way around this difficulty, as it allows one to dispense altogether with electron objectives. Micrographs are obtained in a two-step process, by electronic analysis, followed by optical synthesis, as in Sir Lawrence Bragg's 'X-ray microscope'. But while the 'X-ray microscope' is applicable only in very special cases, where the phases are known beforehand, the new principle provides a complete record of amplitudes and phases in one diagram ... It is a striking property of these diagrams that they constitute records of three-dimensional as well as of plane objects. One plane after another of extended objects can be observed in the microscope, just as if the object were really in position. — D. Gabor.

From *Nature* 15 May 1948.

Planet formation

Grainy pictures of new worlds

Vincent Mannings

We know that planetary bodies in our Solar System began by agglomeration of dust grains, rocks and mountain-sized asteroids. We also know that most of the work was done during the first 10 million years (Myr) after the Sun's birth, when the young star was encircled by an immense disk of grains and gas. But to go beyond this is not easy. Some 4.5 billion years after the main event, examination of our neighbouring worlds may never allow us to reconstruct fully how the Sun's early disk evolved into gaseous behemoths such as Jupiter and small rocky planets such as Earth. We must look elsewhere for better clues to our origins, and we find them in new studies of disks around three stars by Holland *et al.*¹, and of a disk surrounding the star HR4796A by Jura *et al.*², Jayawardhana *et al.*³ and Koerner *et al.*⁴.

Dust grains orbiting a star are bathed in the star's radiation field, and re-radiate this light at millimetre and infrared wavelengths. Holland *et al.*¹ used an array of millimetre-wave detectors to map the distribution of grains in disks around the nearby stars β Pictoris, Fomalhaut and Vega. β Pic is between 10 and 100 Myr old, Fomalhaut is around 200 Myr, and Vega is the oldest at 350 Myr. So with the possible exception of β Pic, each disk is now far beyond the main era of planet-building, and the grains are thought to be debris generated by collisions of asteroids formed earlier on.

All three debris disks have been known of⁵ since the early 1980s, but only the edge-on disk around β Pic has been imaged before⁶. The new millimetre images¹ supply fresh information on each disk. In β Pic we again see the edge-on structure, but there is also a compact source about 700 AU from the star (1 AU is the mean separation of the Earth and the Sun). This object is to the southwest of β Pic. Perhaps it is some sort of disruption in the outer disk. Alternatively, it could be a distinct cloud of grains surrounding a low-mass companion star or sub-stellar object.

Holland *et al.* show that the 'disk' around Fomalhaut is toroidal, with an inner cavity of radius 30 AU. This is the radius of the orbit of Neptune, and it is also comparable to the radius of a cleared-out region of β Pic's disk seen⁷ at mid-infrared wavelengths (0.01 mm, or 10 μ m). Many of the grains formerly at distances less than 30 AU from Fomalhaut may have accumulated during an epoch in which a large population of asteroids, or 'planetesimals', was created within the very dense inner regions of Fomalhaut's primordial disk. These grains,

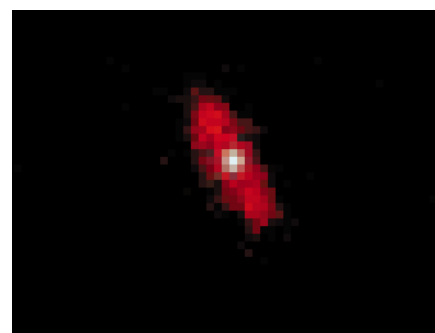


Figure 1 An edge-on disk around the young star HR4796A. The star appears white in this composite infrared image. 13- μ m emission (bluish-green) from warm grains is seen near to the star, and 21- μ m emission (red) is mostly from a cool outer disk.

now locked up inside the planetesimals, are no longer visible at millimetre wavelengths, and so we see a cavity in the debris disk. In the Sun's disk, planetesimals were the primary building blocks⁸ of both terrestrial rocky planets and the cores of gas giants like Jupiter, raising the possibility that the cavity around Fomalhaut may also harbour one or more planets.

The disk surrounding Vega is face-on to us, and there is a bright and elongated central region. This structure is centred roughly on the star and the northeast end is marginally brighter than its southwestern counterpart. What is it? The whole feature could be the result of gravitational perturbations of the grains by an unseen planet. Or maybe the northeastern enhancement is a compact object surrounded by a cloud of grains — a circumplanetary disk?

The star HR4796A is much younger, just 10 Myr old, or about the age at which the Sun's disk was well on its way to converting the bulk of its material into planets. Its environment is mid-way between the massive, pre-planetary gaseous disks orbiting some very young stars^{9,10} and the tenuous dusty debris disks around the older stars just described. Jura *et al.*² show that a cloud of grains also encircles HR4796A, probably arranged in a disk with an inner cavity. The grains have a temperature of about 100 K, implying that they reside at a radius of 35 AU and beyond; the cavity is similar in size to the cleared-out regions around β Pic and Fomalhaut. This is confirmed by Jayawardhana *et al.*³: at 19 μ m they see that the cloud is a disk oriented nearly edge-on, with an outer radius of at least 130 AU. Best of all, they see a direct hint of the inner cavity. Koerner *et al.*⁴ present higher-resolution images at 13 and 21 μ m (Fig. 1). Whereas

the cavity is now seen more clearly at 21 μm , much of the 13- μm emission comes from within this inner 'hole'. Further, there is residual 21- μm emission from the same region. The ratio of the 13- and 21- μm excess fluxes from the inner region implies warm grains at a mean temperature of 250 K. That places them just 3 to 6 AU from the star.

So the HR4796A system appears to begin with a cool outer disk from about 35 to 130 AU. That corresponds to the location of the Kuiper Belt of comets at the outskirts of our own Solar System, as indeed do the locations of the disks around β Pic and Fomalhaut. The grains orbiting far from the young HR4796A may therefore be the precursors and/or debris of comets². The space within 35 AU is not a void, but instead is home to at least a population of warm grains close to the star. We also find grains spread throughout our inner Solar System, primarily debris from collisions of asteroids, many of which reside between about 2 and 5 AU from the Sun. So the warm grains around HR4796A may be the signature of a population of asteroids existing at about the same relative position as most of the asteroids in our Solar System. More speculatively, the presence of asteroids may in turn signal the existence of planets. As for β Pic and maybe Fomalhaut, the very existence of a cavity within 35 AU of HR4796A may itself betray the presence of planets, freshly created from the dense material formerly comprising the inner primordial disk. These planets could then have completed the evacuation of the cavity by gravitationally clearing out their surroundings.

Planets are invoked at almost every turn. Are these four stars really accompanied by planetary families? Certainly there are strong indications that the disks are today being replenished by dust from disruptions of unseen larger bodies. Planets are probably in there, somewhere, but the authors of all these papers are rightly cautious, and we cannot yet make a secure leap from grains to young planetary systems. Although the disks may contain asteroids and grainy debris plus some mix of rocky, Earth-like planets and gaseous giants, they may instead simply be composed of asteroids, debris and nothing else. And what of our understanding of how planets are created? The most decisive way to explore such events is, of course, to observe directly young planets in the act of forming. That is for the future. In the meantime, the superb results described here will surely help us unravel the processes by which circumstellar disks evolve into planetary bodies. □

Vincent Mannings is at the California Institute of Technology, Department of Astronomy, MS 105-24, Pasadena, California 91125, USA.

- Holland, W. S. *et al.* *Nature* **392**, 788–792 (1998).
- Jura, M. *et al.* *Astrophys. J.* (in the press).
- Jayawardhana, R. *et al.* *Astrophys. J.* (submitted).
- Koerner, D. W., Ressler, M. E., Werner, M. W. & Backman, D. E. *Astrophys. J.* (submitted).
- Backman, D. E. & Paresce, F. in *Protostars and Planets III* (eds Levy, E. H. & Lunine, J. I.) 1253–1304 (Univ. Arizona Press, Tucson, 1993).
- Smith, B. A. & Terrile, R. J. *Science* **226**, 1421–1424 (1984).
- Lagage, P. O. & Pantin, E. *Nature* **369**, 628–630 (1994).
- Lissauer, J. J. *Annu. Rev. Astron. Astrophys.* **31**, 129–174 (1993).
- Mannings, V., Koerner, D. W. & Sargent, A. I. *Nature* **388**, 555–557 (1997).
- Beckwith, S. V. W. & Sargent, A. I. in *Protostars and Planets III* (eds Levy, E. H. & Lunine, J. I.) 521–542 (Univ. Arizona Press, Tucson, 1993).

increasing the number of lever presses required for drug delivery. They found that, as the workload increased, the knockout mice were much more willing to continue responding for cocaine, and, on average, they received twice as many injections as the wild-type mice.

Rocha *et al.* next tested whether, during repeated administration of cocaine, the lack of a serotonin-1B receptor would alter the normal process of behavioural sensitization. This term refers to a progressive increase in certain effects of psychostimulants as these drugs are repeatedly administered². Although controversial, it has been argued that addiction results from changes in the brain dopamine system that may be responsible for 'wanting' drugs. In other words, 'wanting' — but not necessarily 'liking' — becomes sensitized³.

The authors placed the mice in an open field, and locomotor activity (running) was detected by a video tracking system. The mice were also watched for the repetitive stereotypic behaviour, such as sniffing, rearing and head-weaving, that characterizes the stimulant effects of cocaine. During the initial exposure to cocaine the knockout mice ran about more than the wild-type mice, and, at relatively low doses of cocaine, stereotypic head-weaving was seen only in the knockout mice. After five daily injections of cocaine, wild-type mice were sensitized to both stereotypic behaviour and locomotor activity, whereas the knockout mice were sensitized only to the running response. These findings indicate that the serotonin-1B receptor null mutation had already produced a degree of sensitization to cocaine, because the knockout mice were more active initially and they showed stereotypic behaviour at what were sub-threshold doses in normal mice.

If the knockout mice were, in fact, pre-sensitized to cocaine, they might also show the neurochemical alterations that are known to be associated with cocaine sensitization. These include increased induction of a set of proteins that are related to the immediate early gene *c-fos*. Termed chronic Fos-related antigens (FRAs) because they are induced specifically by chronic treatment with cocaine and other drugs⁴, they correspond to post-transcriptional modifications of ΔFosB ⁵, a truncated splice variant of FosB⁶. Chronic FRAs interact with products of another immediate early gene, *c-jun*, to form a DNA-binding complex known as activator protein-1 (AP-1) that regulates the transcription of target genes. Because chronic FRAs are expressed only with chronic cocaine treatment, and expression persists long after the cocaine has been withdrawn, this molecular adaptation is thought to be closely associated with the processes that underlie addiction to cocaine⁷.

Drug addiction

Cocaine and the serotonin saga

Francis J. White

Cocaine addiction remains a serious social problem, largely because we lack effective treatments for it. We know that, in the brain, the main molecular targets of cocaine are the transporter proteins that remove the neurotransmitters dopamine, noradrenaline and serotonin (5-hydroxytryptamine) from synapses after they have sent their messages to postsynaptic neurons. We also know that projection of the dopamine system to a region of the forebrain called the nucleus accumbens is mainly involved in the euphoric (rewarding) and psychostimulant effects of cocaine. Yet, paradoxically, cocaine has a higher affinity for the serotonin transporter than for the dopamine transporter.

On page 175 of this issue, Rocha *et al.*¹ have used mice that lack the serotonin-1B receptor to provide the first compelling evidence that a serotonin receptor is involved

in the processes that underlie vulnerability to cocaine addiction. Although there had been suggestions that serotonin modulates cocaine-induced euphoria in a negative manner, the data were subject to different interpretations. Moreover, the diversity of serotonin receptors (at least 14), and the lack of ligands with necessary selectivity for the various receptors, made pharmacological analysis particularly difficult. The availability of gene 'knockout' mice now provides a means to circumvent these problems.

Rocha and colleagues determined the 'willingness' of serotonin-1B receptor knockout mice and their wild-type littermates to work for intravenous injections of cocaine. The mice were initially trained to self-administer cocaine by pressing a lever — each press resulted in a single shot of cocaine. Once the mice were trained, the authors 'upped the ante' by gradually

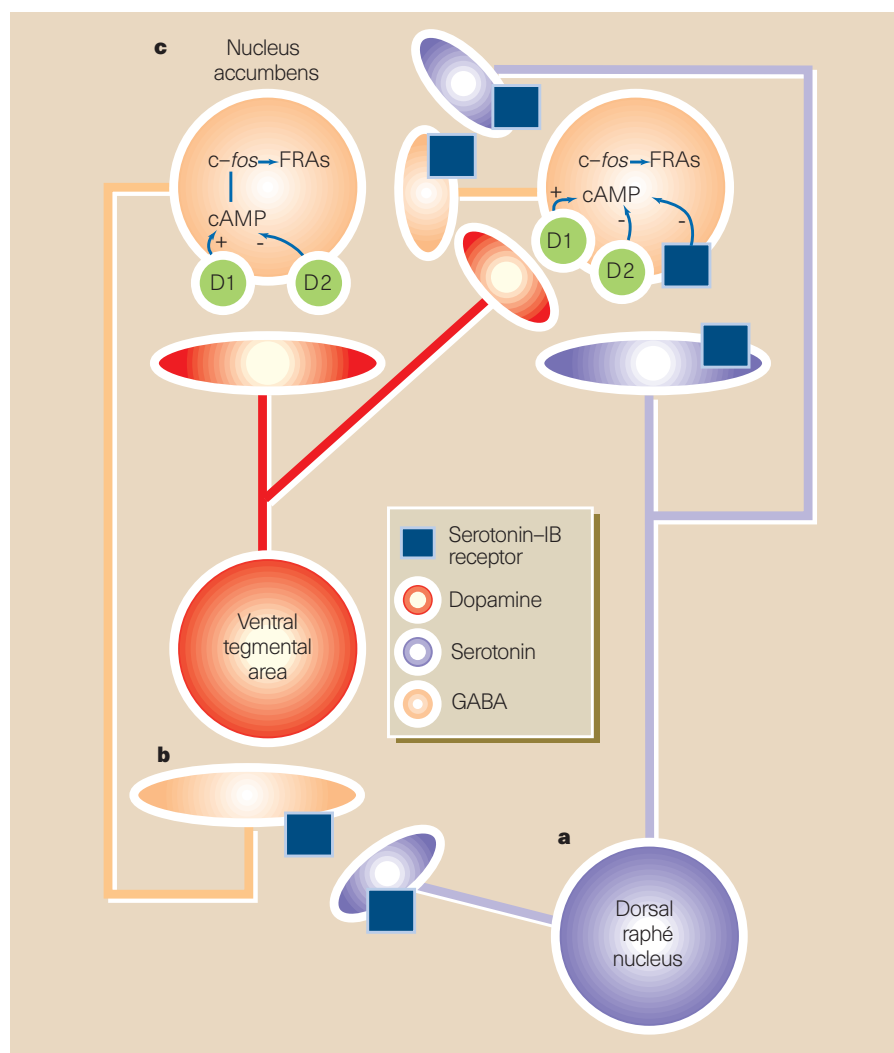


Figure 1 Possible sites at which function of the dopamine system might be affected by loss of serotonin-1B receptors. By studying mice lacking the serotonin-1B receptor, Rocha *et al.*¹ have found that this receptor is involved in the processes that underlie addiction to cocaine. **a**, Serotonin-releasing neurons originate in the dorsal raphe nucleus and project to both the ventral tegmental area (VTA) and the nucleus accumbens (NAc). The serotonin-1B receptors act as autoreceptors on serotonin-releasing neurons, regulating this release. If this population of receptors is lost, serotonin transmission may be increased in both the VTA and the NAc. **b**, Serotonin-1B receptors on the terminals of γ -aminobutyric acid (GABA)-releasing neurons within the VTA suppress release of GABA. Loss of this modulation in the knockout mice could lead to activation of dopamine-releasing neurons, which, in turn, would increase stimulation of dopamine-D1 receptors, leading to activation of *c-fos* and production of Fos-related antigens (FRAs). Serotonin-1B receptors might also modulate GABA transmission within the NAc. **c**, Somatodendritic regions of NAc neurons may express serotonin-1B receptors, which are negatively coupled to adenyl cyclase and the cyclic AMP pathway. Loss of these receptors in the knockout mice might increase signalling through the dopamine-D1 receptor, resulting in the production of chronic FRAs.

Rocha and colleagues found that several Fos-related proteins — including the chronic FRAs — were more abundant in extracts from the nucleus accumbens (and a closely related dorsal striatal area) in the knockout mice than in the wild-type controls. Moreover, in the knockout mice, more cells in the nucleus accumbens were immunoreactive against an anti-chronic FRA antibody. Finally, there was considerably more of the AP-1 transcription complex in the striatum and nucleus accumbens of the knockouts. This is consistent with

dimerization of the induced chronic FRAs with Jun-related transcription factors. To tie these molecular alterations back to cocaine addiction, the authors showed that by giving the wild-type mice chronic cocaine, the amount of the AP-1 transcription complex became comparable to levels in untreated knockout mice.

These results provide the first definitive evidence for the involvement of a specific serotonin receptor in processes that may underlie cocaine addiction. They also raise mechanistic questions. For example, do

serotonin-1B receptors normally act as a brake on the brain's 'want' system, such that their removal leads to a sensitized state similar to cocaine addiction? If so, does this imply that chronic levels of cocaine down-regulate the signalling system to which the serotonin-1B receptors belong in normal mice (and humans)? And how do the serotonin-1B receptors normally modulate chronic FRAs? Serotonin-1B receptors are found mainly at nerve terminals, including the normally inhibitory synaptic connections to dopamine neurons that innervate the nucleus accumbens and dorsal striatum, and the authors suggest that overactivity of these dopamine neurons may cause the changes observed in the knockout mice (Fig. 1).

As Rocha *et al.* point out, their findings seem to contradict previous studies that used ligands selective for the serotonin-1B receptor to modify the behavioural and neurochemical effects of cocaine. Results from those experiments indicated that certain effects of cocaine could be mimicked by stimulating the serotonin-1B receptors, and that blocking these receptors might reduce the effects of cocaine. Why, then, would removal of the serotonin-1B receptors result in sensitized cocaine responses and cocaine-like cellular adaptations? Rocha *et al.* suggest that the knockout mice might have developed compensatory mechanisms to produce the observed behavioural and cellular phenotypes. As well as exploring this possibility, we need to test whether the previous pharmacological analyses might have been compromised by the lack of receptor selectivity of many drugs used in such experiments.

Whatever the case, the elegant work of Rocha and co-workers, along with studies of other serotonin-receptor knockout mice (such as the serotonin-2C receptor⁸), should allow us to address the question of how serotonin is involved in cocaine addiction with greater precision than ever. The importance of this is underscored by additional findings by Rocha *et al.*⁹, to be reported in the June issue of *Nature Neuroscience*, whereby mice that lack the dopamine transporter will self-administer cocaine. □

Francis J. White is in the Department of Cellular and Molecular Pharmacology, Finch University of Health Sciences/Chicago Medical School, 3333 Green Bay Road, North Chicago, Illinois 60064-3095, USA.

e-mail: whitef@mis.finchcms.edu

1. Rocha, B. A. *et al.* *Nature* **393**, 175–178 (1998).
2. Kalivas, P. W. & Stewart, J. *Brain Res. Rev.* **16**, 223–224 (1991).
3. Robinson, T. E. & Berridge, K. C. *Brain Res. Rev.* **18**, 247–291 (1993).
4. Hope, B. T. *et al.* *Neuron* **13**, 1235–1244 (1994).
5. Hiroi, N. *et al.* *Proc. Natl Acad. Sci. USA* **94**, 10397–10402 (1997).
6. Nakabeppu, Y. & Nathans, D. *Cell* **64**, 751–759 (1991).
7. Nestler, E. J. *Curr. Opin. Neurobiol.* **7**, 713–719 (1997).
8. Tecott, L. H. *et al.* *Nature* **374**, 542–546 (1995).
9. Rocha, B. A. *et al.* *Nature Neurosci.* **1**, 132–137 (1998).

Neutron stars

How pulsars get their spins

Adam Burrows

A supernova explosion announces the spectacular death of a star heavier than about eight solar masses. Its brightness can rival that of its host galaxy, and its total explosion energy is a prodigious 10^{51} ergs. The usual residue of such an explosion, a neutron star, is no less extreme an animal. With a density that of the atomic nucleus, these stars the size of a city are observed to be rotating with periods from milliseconds to seconds, and travelling at up to $1,500 \text{ km s}^{-1}$ (Fig. 1, overleaf). Radio pulsars are neutron stars, and as such have been studied for 30 years. Approximately 10^3 of probably 10^5 radio pulsars in the Galaxy have been discovered (among perhaps 10^8 neutron stars). Their masses, magnetic fields, spin periods, spin-down rates, and proper motions across the sky are catalogued, and theories for their emission are devised. But the origin of these basic pulsar parameters has all but been ignored as untestable. On page 139 of this issue¹, Spruit and Phinney attempt to bridge the gap between the supernova and pulsar communities with a provocative thesis.

Taking a lead from the suggestion² that a

'kick' from an asymmetrical supernova could exert a torque on a nascent neutron star, they posit a unified theory of pulsar spins, kicks and magnetic fields. First, they suggest that magnetic fields in massive stars couple core and mantle strongly enough to leave the pre-supernova core in solid-body rotation. Such initial magnetic fields indeed yield neutron stars with fields of 10^{12} gauss, as seen in radio pulsars. But the core would rotate far too slowly (with a period of ~ 100 seconds) to leave young neutron stars spinning as fast as they do. So Spruit and Phinney propose that the mechanism that gives the neutron star its linear kick may also generate its spin.

Second, the authors show that a series of random impulses naturally yields a quantitative connection between spin and kick magnitudes, and at the same time partially decorrelates spin and transverse proper motion vectors and magnitudes. A single kick would produce a spin perpendicular to the velocity change, and so give a strong correlation between spin and apparent motion on the sky — which radio pulsar data do not show. Many randomly placed kicks, perhaps from a convecting supernova explosion,

would not cause the same problem.

Evidence that neutron stars experience a net kick at birth has been mounting for years. Pulsars are the fastest population in the Galaxy³, with an average velocity of $\sim 450 \text{ km s}^{-1}$. Such speeds are far larger than any orbital motion left over from birth in a binary. In the pulsar binaries, PSR J0045–7319 and PSR1913+16, the spin axes and the orbital axes are misaligned, suggesting that the explosions that created the pulsars were not spherical (ref. 4, and I. Wasserman, J. Cordes and D. Chernoff, personal communication). In fact, for the former the orbital motion seems retrograde relative to the spin⁵ — the explosion may have kicked the pulsar backwards! The young supernova remnant Cassiopeia A, formed in about AD 1680, has calcium, sulphur and oxygen element distributions that are clumped and have gross back-front asymmetry⁶ — no simple shells are seen. Many other supernova remnants, such as N132D and SN0540–69.3, have velocities⁷ relative to the local interstellar medium of up to 900 km s^{-1} . The supernova SN1987A is a case study in asphericity: its X-ray, gamma-ray and optical fluxes and light curves require that shards of the radioactive isotope ^{56}Ni were flung far from the core in which they were created; the infrared line profiles of its oxygen, iron, cobalt, nickel and hydrogen are ragged and show a pronounced red–blue asymmetry; its light is polarized (something a sphere cannot accomplish); and Hubble Space Telescope pictures of its inner debris⁸ reveal large clumps and hint at a preferred expansion direction.

Finally, SN1997X is one of the most intriguing recent finds. This is a so-called type-Ic supernova, thought to be the explosion of the bare carbon/oxygen core of a massive star that has been stripped of its envelope. SN1997X has the greatest optical polarization of any known supernova (L. Wang, personal communication), which implies that supernova cores, and explosions, are asymmetrical.

All in all, it seems clear that neutron stars can be given a hefty extra kick at birth, and that asymmetries in the supernova explosion are implicated. How might these asymmetries arise? No doubt, instabilities in the outer envelopes of supernova progenitors make debris clouds clumpy and mixed. The observation of hydrogen deep in SN1987A's ejecta strongly suggests the work of such mantle instabilities. But overall, the data — particularly for the heavier elements produced in the inner core — point to asymmetries in the central engine of the explosion.

Supernova theorists have not been napping, and over the past decade have determined that supernova cores are indeed grossly unstable to convective instabilities^{2,9,10}. Between the 'bounce' (when some material falls in and effectively bounces off

Entomology

Beauty and the beetle

Coloured beetles are of two basic types. Some have a layer of pigment in their exoskeleton (pigmentary colours), while others rely on the interaction of light with features in their integument (structural colours). These structurally reflecting colours can be very different — *Calloodes grayanus* (pictured right), for example, appears a weak green in colour, whereas *Anoplognathus parvulus* is a strong metallic gold. How are these different colours produced?

Writing in the *Journal of Experimental Biology* (201, 1307–1313; 1998), Andrew R. Parker and colleagues provide the answer. In both cases, the structural colour arises from so-called 'multilayer reflectors' — layers of a transparent material separated by layers with a different refractive index. Reflected beams from all of the layers interfere, and when this interference is constructive for a particular wavelength of light, that colour is observed.

But the similarity ends there. Parker *et al.* have found that the *C. grayanus* reflector is composed of regularly spaced layers of the same optical thickness, overlaid by an irregular transparent



layer that scatters light. The resulting reflection is diffuse, appearing the same when viewed from any direction; this green colour seems to provide excellent camouflage against a leafy background.

The layers that make up the *A. parvulus* reflector, however, decrease in spacing (and consequently in optical thickness) with depth in the structure. The resulting metallic gold colour can be seen only from certain directions, but it makes *A. parvulus* very conspicuous nonetheless, perhaps allowing it to be recognized by members of the same species. Unfortunately, its striking vesture could also mark it as a tasty morsel for predators.

Alison Mitchell

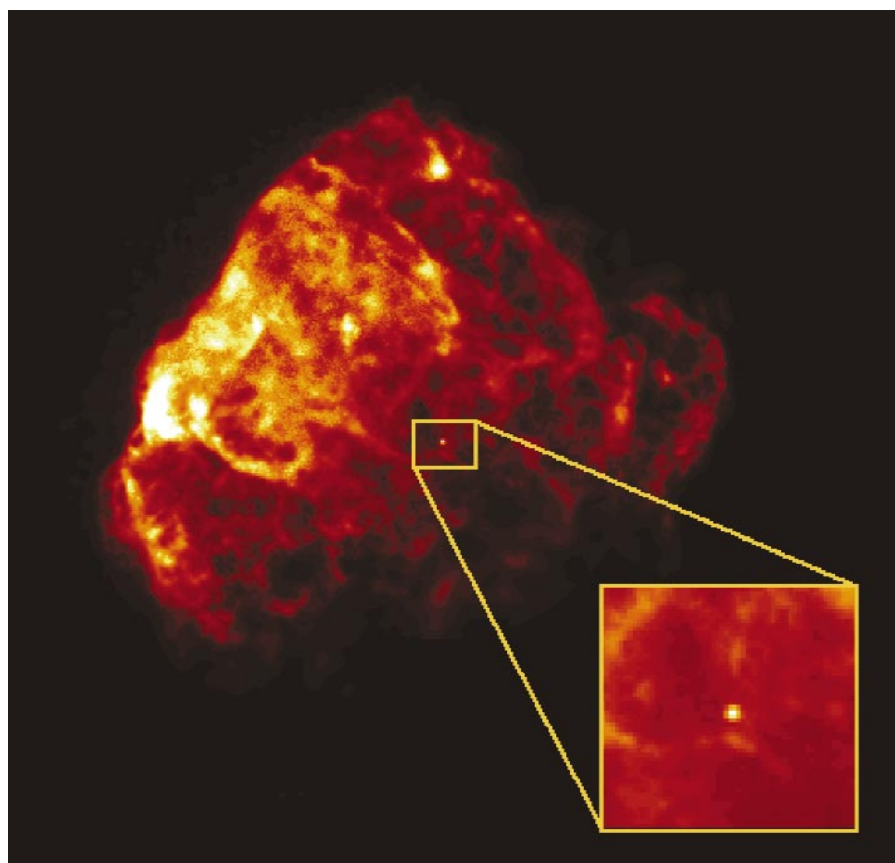


Figure 1 The 3,700-year-old supernova remnant Puppis A, imaged in X-rays by the satellite Rosat. The bright spot has been interpreted as its neutron star, even though no pulsations are seen, because it has a large ratio of X-ray to optical emission. Its inferred transverse speed is $\sim 1,000 \text{ km s}^{-1}$. Interestingly, the spot is on the opposite side to the fast, oxygen-rich knots, as one might expect in some models of neutron-star recoil during a supernova explosion.

the surface of the forming neutron star) and the explosion, there is a delay while the outward-moving shock wave travels through the star. During this 100 to 1,000 milliseconds, the core is strongly convective, boiling and churning at sonic speeds ($\sim 3 \times 10^4 \text{ km s}^{-1}$). Any slight asymmetry in collapse can amplify this jostling, and result in vigorous kicks and torques to the residue that can be either systematic or random¹¹.

Whatever the details, it would seem odd if the nascent neutron star were not left with a net recoil and spin, although it is not known whether pulsar speeds as high as $1,500 \text{ km s}^{-1}$ can be reached through this mechanism. But asymmetries in the matter distribution may also cause asymmetries in the emission of neutrinos, which carry away most of the binding energy of the neutron star (around 3×10^{53} ergs). Amazingly, a net angular asymmetry in the neutrino radiation of only 1% would give the residue a recoil of 300 km s^{-1} . Not surprisingly, many theorists have concentrated on producing such a neutrino asymmetry, invoking anisotropic accretion, exotic neutrino-flavour physics, or the influence of strong magnetic fields on neutrino cross-sections and transport. The last is particularly interesting, but generally requires¹² magnetic fields of 10^{14} to 10^{16} gauss,

far larger than the average pulsar surface field of 10^{12} gauss. Perhaps the pre-explosion convective motions themselves can generate the required fields by dynamo action.

It is now reasonable to imagine a theory that unifies the spins and velocities of neutron stars, the anisotropies observed in supernova ejecta, and stellar collapse and explosion. To date, dialogue between the pulsar and the supernova communities has been rare — but the paper by Spruit and Phinney is a good conversational gambit. □

Adam Burrows is in the Department of Astronomy, University of Arizona, Tucson, Arizona 85721, USA.
e-mail: burrows@jupiter.as.arizona.edu

1. Spruit, H. & Phinney, E. S. *Nature* **393**, 139–141 (1998).
2. Burrows, A., Hayes, J. & Fryxell, B. A. *Astrophys. J.* **450**, 830–850 (1995).
3. Lyne, A. & Lorimer, D. R. *Nature* **369**, 127–129 (1994).
4. Kaspi, V. M., Bailes, M., Manchester, R. N., Stappers, B. W. & Bell, J. F. *Nature* **381**, 584–586 (1996).
5. Lai, D., Bildsten, L. & Kaspi, V. M. *Astrophys. J.* **452**, 819–824 (1995).
6. Lawrence, S. S. *et al. Astron. J.* **109**, 2635–2651 (1995).
7. Kirshner, R. P., Morse, J. A., Winkler, P. F. & Blair, J. P. *Astrophys. J.* **342**, 260 (1989).
8. Pun, C. S. J., Kirshner, R. P., Garnavich, P. M. & Challis, P. *Bull. Am. Astron. Soc.* **191**, 9901 (1998).
9. Herant, M., Benz, W., Hix, J., Fryer, C. & Colgate, S. A. *Astrophys. J.* **435**, 339 (1994).
10. Janka, H.-T. & Müller, E. *Astron. Astrophys.* **290**, 496–502 (1994).
11. Burrows, A. & Hayes, J. *Phys. Rev. Lett.* **76**, 352–355 (1996).
12. Lai, D. & Qian, Y.-Z. *Astrophys. J.* (in the press).

Daedalus

Deceptive appearances

Few women have ever been satisfied with their figures. Hence the foundation-garment industry. Perhaps its peak was the late nineteenth century, when the fashionable Victorian young lady seems to have carried nearly as much rigging as a sailing ship.

Since then, elasticated undergarments have grown simpler and less fashionable. Modern women are urged to perfect their figures by diet and exercise. This state of affairs is not an improvement — elastic at least worked. So Daedalus is studying the basics of the technology.

Tensioned clothing cannot truly shape; it can only squeeze. The natural resulting cross-sections are circles or arcs of circles. The human figure, however, tends to have elliptical or ovoid cross-sections. To impose such shapes, a garment needs not tightness, but directional rigidity.

So DREADCO corsetières are devising fabrics woven, not of fibres which can only carry tension, but of micro-girders which can resist bending. These tiny I-beams have polyimide compression faces, carbon fibre tension faces, and thermoplastic central webs. As with some steel girders, they are initially made in T-section, with a sawtooth vertical web. When two such sections are welded together by their tooth points, the resulting I-beam has diamond-shaped holes along its web. The warp and weft half-girders of the new fabric are laid on a shaped former, the corresponding top half-girders are brought down on them, and all the welds are made at once by a fast-set adhesive. Warp and weft micro-girders thus pass through each other's web holes. They can move angularly against each other, giving the new fabric the soft 'drape' of cloth. But it has a definite shape derived from its former; and will impose that shape, subtly but firmly, on the wearer.

Until the complex micro-engineering of production has been optimized, DREADCO's 'Shapers' will be vastly expensive. In the fashion trade, fortunately, this is not necessarily a disadvantage, especially for such a revolutionary product. Instead of merely squeezing the wearer at one point and displacing the tissue elsewhere, they will truly shape her figure. Thin and seemingly insubstantial, they will flatten an excessive stomach without cramping the rest of the waist, and will form thighs, buttocks and breasts into pert, becoming contours without even feeling tight. Even if quite overweight, the happy wearer will still display a perfect figure, but slightly scaled up.

David Jones

Kepler's cosmos

Copernicus's system of the Universe was revolutionary but his method of representing it on paper was anything but. It was left to Kepler to apply Renaissance techniques of spatial visualization to make the theory come alive.

Martin Kemp

Nicolaus Copernicus's programme to purge the Ptolemaic model of the Universe of its growing burden of disfiguring complications was driven not least by what we would call aesthetic considerations. The architecture of his heliocentric system reinstated the geometrical integrity that the Greek astronomers had sought: "At rest... in the middle of everything is the sun. For in this most beautiful temple, who would place this lamp in another or better position than that from which it can light up the whole thing at the same time?"

Although Copernicus's "temple" obeyed the principles of harmonically unified design advocated in the Renaissance, his conventionally diagrammatic representation of his scheme in 1543 as a flat series of concentric circles did not avail itself of the new spatial vision inherent in the buildings and paintings of the Renaissance masters. It was more than 50 years later that Johannes Kepler, fervent Copernican and Platonist, allied the new visual forms with the new astronomical vision.

Among the many testimonies to Kepler's extraordinary powers of spatial visualization, none is more remarkable than the great cosmological model he illustrated in a fold-out plate in his *Mysterium Cosmographicum* of 1596. We know how the scheme came to be envisaged. He tells how, when he was teaching his "students the way Great Conjunctions jump eight signs at a time", he drew "many triangles, or quasi triangles, in the same circle".

It began to dawn on him that the distances between the orbits of the planets could be explained if the ratios between successive orbits were designed to be equivalent to the spheres successively circumscribed around and inscribed within the five 'Platonic solids' (the regular polyhedra), nested one inside the other.

Kepler envisaged that his Platonic machine, in its most ambitious form, would be driven by clockwork and that its hollow armature would be filled by alcoholic beverages of a suitable kind. It was, after all, a courtly creation, dedicated to Duke Frederick of Württemberg, who literally acted as a protector for Kepler as a Copernican 'heretic'. The form of visualization and representation in the model stood in direct line of succession from the form of spatial illustration invented by Leonardo da Vinci for

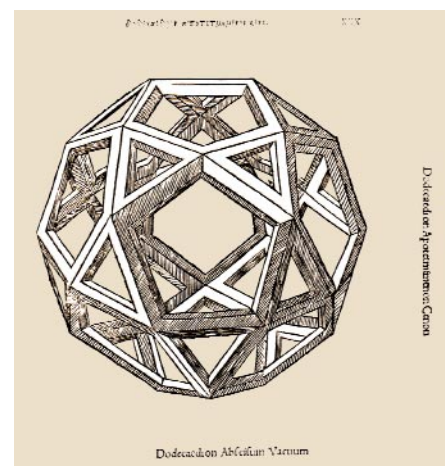


Kepler's "Model of the Orbits of the Planets" from *Mysterium Cosmographicum*, 1596 (above).

Woodcut after Leonardo da Vinci, "Truncated Dodecahedron", from Luca Pacioli's *De Divina Proportione*, 1509 (right).

the treatise on the five regular polyhedra, *De Divina Proportione*, originally written in 1498 by his colleague in Milan, Luca Pacioli.

For Kepler, the pervasiveness of mathematical harmonics was such that they not only explained the music of the heavens, but also provided a model for forms of government. Taking his cue from the French philosopher Jean Bodin's *Six Books of the Republic* (1576), he set out broad equations between aristocratic republics and the kind of proportions expressible in numbers, and between royal governance and the incommensurable properties of geometrical proportions. Kepler argued that the harmony of proportion should ideally permeate every relationship in society from the formulation of laws to the weighting of punishment and from the



delights of friendship to the division of inheritances. Tempered proportion and apportionment is the rule in all things. The aim is that things should be on Earth as in the heavens. □

Martin Kemp is in the Department of the History of Art, University of Oxford, 35 Beaumont Street, Oxford OX1 2PG, UK.

e-mail: martin.kemp@trinity.oxford.ac.uk

Imprinted gene in postnatal growth role

Mice that have been specially bred to lack a protein known as Grf-1, which is normally found only in the brain, do not grow properly after they are born and remain small all their lives. We have now identified a function of Grf-1 as an important regulator of the synthesis and release of growth hormone. Moreover, *grf1*, the gene encoding this protein, is unlike other imprinted genes that affect growth because it operates after, rather than before, birth.

Grf-1 is a brain-specific factor¹ of relative molecular mass 140K that promotes the exchange of GDP for GTP on membrane-bound Ras proteins, thereby regenerating the active form of these proteins². Ras proteins are guanine-nucleotide-binding proteins (G proteins) that are involved in the transduction of signals controlling cell growth and differentiation³. Here we investigate the effects of Grf-1 on mouse development by disabling the *grf1* gene.

We disrupted the *grf1* allele by gene targeting in embryonic stem cells. A deletion

in an exon coding for the catalytic domain inactivated all of the potential splicing isoforms⁴. Heterozygous and homozygous mutant *grf1* mice were obtained in the predicted mendelian ratio, and were viable and fertile in both the pure 129 sv and mixed (129 sv × C57black6/Jlco) backgrounds.

In the brains of adult wild-type mice, but not in other tissues tested (data not shown), a 5.5-kilobase transcript and two proteins of 140K and 130K were detected using a monoclonal antibody raised against the catalytic domain of Grf-1 (Fig. 1a, b). In homozygous mutant mice, only the 130K protein is present and it is unaltered; the 5.5-kilobase RNA and 140K protein are not evident (Fig. 1a, b). It is likely that the 130K protein is the recently identified *grf2* gene product as the two proteins share 80% homology in their catalytic domain⁵. Analysis of heterozygous mutant mice confirmed that *grf1* is an imprinted gene in mice⁶: only the paternal allele is expressed (Fig. 1a, b).

During embryogenesis and immediately

Table 1 Reduced growth hormone levels in pituitaries of mutant mice¹

Sex	No. of mice +/+	No. of mice +/-	Age (days)	Pituitary GH reduction in mutant mice
M	3	3	14–17	~42%
F	2	2	15	~53%
F	2	2	31	~46%

¹The concentration of growth hormone was determined by radioimmunoassay. GH, growth hormone; +/+, wild-type mice; +/-, mutant mice.

after birth, mutant and wild-type mice showed no apparent difference in their development (Fig. 1c). This accords with the timing of Grf-1 expression — the protein is not detected in the fetus, is barely detectable at birth, but is clearly visible on the second postnatal day (Fig. 1b). Postnatally, the body weight of wild-type and heterozygous +/- (maternal/paternal allele) mice increased more rapidly than that of heterozygous +/- and homozygous -/- mutant mice in both backgrounds. The postnatal growth deficiency was most prominent around weaning (3–4 weeks), when body weight was 25 and 30% lower in female and male transgenic +/- mice, respectively. The growth deficiency persisted in adult mice, which were approximately 15–25% smaller (Fig. 1c,d). This phenotype was similar in homozygous -/- mice of both sexes.

We tested for the expression of growth hormone and insulin-like growth factor-1 (Igf-1) in these *grf1*-deficient animals and found that, at 4 weeks old, the Igf-1 level in serum was reduced by 32–40% in females and males, respectively, relative to wild-type animals (data not shown).

Expression of Igf-1 is under the control of circulating growth hormone, whose levels in plasma are subject to circadian variation and very sensitive to external stress factors. We therefore determined the amount of growth hormone in the pituitary glands of 14-, 15-, 17- and 31-day-old *grf1* mutant and wild-type mice of the same litter, and found that in *grf1*-deficient +/- mice these contained 42–53% less growth hormone than the wild-type controls (Table 1). The impact of disrupting *grf1* on the production of growth hormone and Igf-1 must be specific because the concentrations of other growth-controlling hormones, such as thyroid-stimulating hormone and T3/T4 thyroid hormones, were unaffected in the pituitaries and plasma of mutant mice.

Grf-1 is not detected in the pituitary, but is present after birth in the hypothalamus of wild-type mice (Fig. 1e). The synthesis of growth hormone is regulated indirectly by a hypothalamic neuronal network using neuropeptides such as cholinergic and adrenergic neurotransmitters, opioid peptides and neuropeptide Y, which exert their

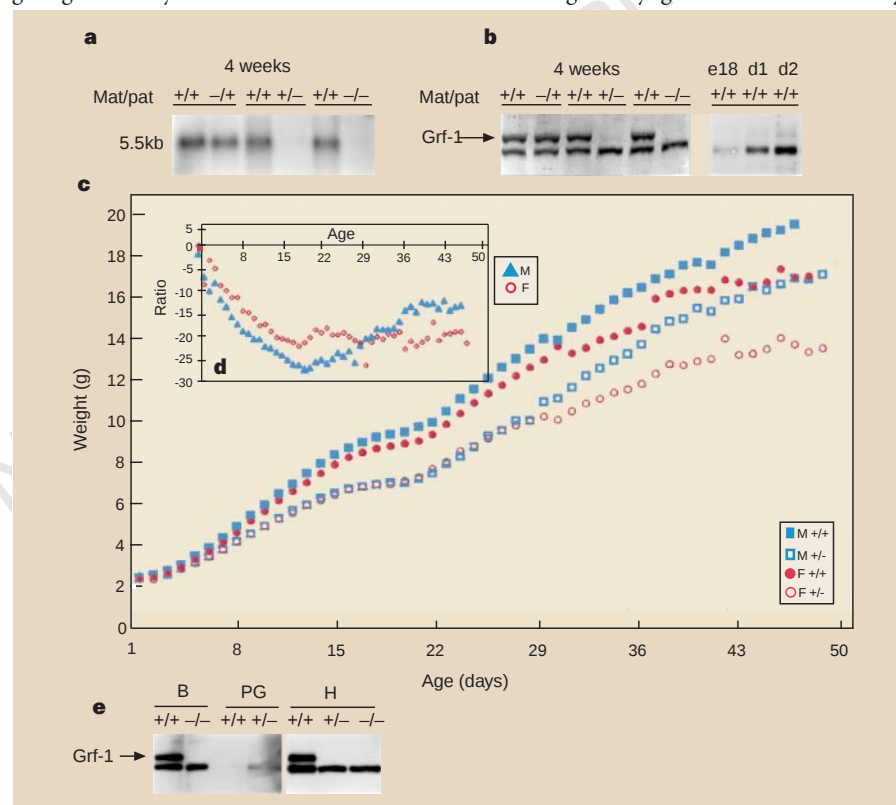


Figure 1 Mouse *grf1* expression and postnatal growth curve **a**, Detection of 5.5-kb *grf1*-specific transcript in brain by northern blot analysis, using 2 μ g poly(A)⁺ RNA from 4-week-olds. None was detected in homozygotes for the mutation (-/-). The paternal allele was fully active in heterozygous +/- mutants; the maternal allele was weakly expressed (<0.1%) in heterozygous +/- mutants. **b**, Detection of Grf-1 in brain. It was absent in homozygotes (-/-) and in heterozygous +/- 4-week-olds. Only the paternal allele of *grf1* was expressed (-/+ lane). It is not detected at embryonic day 18 (e18) but becomes visible by postnatal day 2 (d2). **c**, Mean of postnatal growth of male and female controls (+/+) or mutant (+/-) 129sv mice. Animals were weighed daily from days 1 to 30 and then once a week. Numbers of animals: wild-type males, 50; wild-type females, 55; mutant males, 45; mutant females, 38. **d**, Growth deficiency calculated from curves in **c**. **e**, Detection of Grf-1 in pituitary glands (PG) of 2-month-olds and in hypothalamus (H) of 20-day-olds. B, brain.

effects mainly through hypothalamic growth-hormone-releasing hormone (GHRH) and somatostatin-releasing-inhibiting factor (SRIF)⁷. Perhaps Grf-1-dependent regulation of growth hormone synthesis and release is modified in our *grf1* mutant mice by deregulation of the neuronal network in the hypothalamus. The molecular events that link muscarinic receptors and Ras proteins through Grf-1 (ref. 8) might also be involved in controlling the GHRH/SRIF balance in the hypothalamus and in memory consolidation in the amygdala⁹.

Like other paternally expressed genes, *grf1* is involved in growth stimulation, whereas maternally expressed genes are responsible for growth suppression¹⁰. In contrast to these imprinted genes that are implicated in fetal growth, *grf1* is the first imprinted gene to be implicated exclusively in postnatal growth control.

Jean-Michel Itier, Günter L. Tremp, Jean-François Léonard, Marie-Christine Multon, Grénaëlle Ret, Fabien Schweighoffer*, Bruno Tocqué*

Rhône-Poulenc Rorer, Centre de Recherche de Vitry-Alfortville, Vitry-sur-Seine, France

*Present address: BioScreen Therapeutics, 58 Bd St Denis, 92400 Courbevoie, France

Marie-Thérèse Bluet-Pajot

Biodynamique des Systèmes Neuroendocriniens, INSERM U-159, Centre Paul Broca, Paris, France
Valérie Cormier

Unité de Recherches sur les Handicaps Génétiques de l'Enfant, INSERM U-393, Hôpital des Enfants-Malades, Paris, France

François Dautry

Génétique Moléculaire et Intégration des Fonctions Cellulaires, CNRS UPR 9044, Institut de Recherches sur le Cancer, Villejuif, France

1. Zippel, R. *et al.* *Mol. Brain Res.* **48**, 140–144 (1997).
2. Feig, L. A. *Curr. Opin. Cell Biol.* **6**, 204–211 (1994).
3. Lowy, D. & Willumsen, B. *Annu. Rev. Biochem.* **62**, 851–891 (1993).
4. Cen, H., Papageorge, A. G., Kippel, R., Lowy, D. R. & Zhang, K. *EMBO J.* **11**, 4001–4015 (1992).
5. Fam, N. P. *et al.* *Mol. Cell Biol.* **17**, 1396–1406 (1997).
6. Plass, C. *et al.* *Nature Genet.* **14**, 106–109 (1996).
7. Bertherat, J., Bluet-Pajot, M. & Epelbaum, J. *Eur. J. Endocrinol.* **132**, 12–24 (1995).
8. Mattingly, R. R. & Macara, I. G. *Nature* **382**, 268–272 (1996).
9. Brambilla, R. *et al.* *Nature* **390**, 281–286 (1997).
10. Jaenisch, R. *Trends Genet.* **13**, 323–329 (1997).

Miocene/Pliocene shift: one step or several?

Cerling *et al.*¹ provide evidence of a world-wide expansion of biomass of plants using the C₄ photosynthetic pathway about 8–6 million years (Myr) ago, recorded in the carbon isotope composition of the dental enamel of fossil herbivores. The authors claim that the expansion of C₄ plants is accompanied by a “worldwide faunal change” and infer that “an important global

ecological change was under way at this time”. For these statements to be true, the period of transition from C₃ to C₄ plants must prove to be contemporary with the period of faunal turnover, and both events must have taken place in the same geographic areas. At least, this is what the correlation of faunal and vegetation events provided by Cerling *et al.* seems to indicate.

But this model lacks a detailed analysis of extinction/immigration rates in the framework of precise palaeomagnetic and biostratigraphical data. In this context, a comparison of the late Miocene–early Pliocene records of artiodactyl and rodent communities from Spain and Pakistan is of special interest. First, the amount of palaeomagnetic and biostratigraphical data from both areas^{2–8} provides a high resolution of the shifts in faunal communities, and second, according to Cerling *et al.*¹, the C₃/C₄ transition took place all over the world except in western Europe.

Our comparison of extinction and immigration rates from sites in Spain and Pakistan (Fig. 1) reveals that first, in Pakistan's Siwalik sediments, not one but two faunal changes occurred, affecting especially artiodactyls. The first one, at about 10–9 Myr ago, largely preceded the C₃/C₄ transition (8–6 Myr ago); the second one occurred at its end (6.5 Myr). However, none of them exactly coincides with the initial expansion of C₄ biomass (8–7 Myr ago), despite a slight increase in rodent extinctions.

Second, both faunal changes in the Siwaliks are contemporary with faunal turnovers in Spain, although in western Europe there is no indication of C₄ diet at any time. In Spain, these faunal changes turn out to have been even more dramatic than in the Siwaliks as the extinction/immigration rate is twice as high. In Spain, woodland or forest indicators (hominoids and tragulids) became extinct about 9 Myr ago, whereas in the Siwaliks hominoids survived their European counterparts by about 1 Myr, and tragulids by as much as 5 Myr.

In the light of these data, we cannot find synchronicity or even a causal link between faunal and vegetation change. Faunal turnovers of different magnitudes took place before and after the appearance of C₄ plants, suggesting that C₃/C₄ transition and faunal turnovers occurred independently.

The “important global ecological change” instead looks like a succession of faunal and vegetation changes, scattered over a long period of at least 5 Myr. We feel that there is insufficient information to permit a clear interpretation of these phenomena. They probably occurred under the influence of a variety of events such as alpine orogenesis with Himalayan and Tibetan uplift, monsoonal dynamics, increasing seasonality, and other factors not yet fully investigated.

Meike Köhler, Salvador Moyà-Solà,

Jordi Agustí

Institut de Paleontologia M. Crusafont,

Escola Industrial 23, 08201 Sabadell, Barcelona, Spain

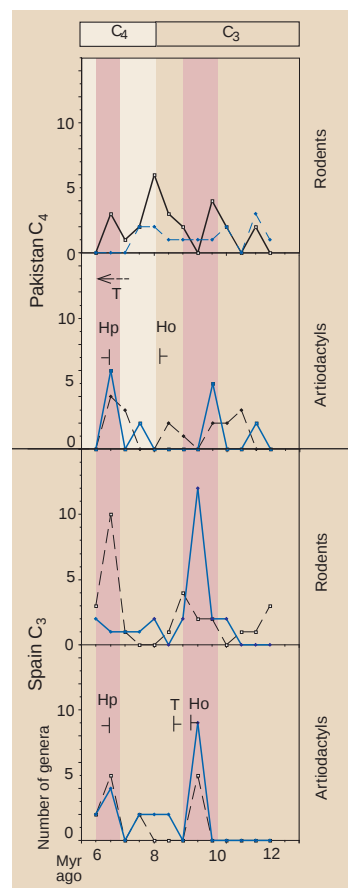


Figure 1 Faunal turnovers in the Eurasian Upper Miocene epoch under C₃ (Spain) and C₄ (Pakistan) conditions. Dotted lines, appearance of genera; continuous lines, disappearance of genera. Hp, hippopotamids; Ho, hominoids; T, tragulids. Data on Pakistan's Siwalik sediments derive from ref. 8; the palaeomagnetic scale is from ref. 9. Considered are herbivore communities with frequent changes at the generic level (artiodactyls and rodents); excluded are groups with low generic diversity such as proboscideans, rhinos and equids. Between 12 and 5.5 Myr ago, two faunal turnovers took place, the older one between 10 and 9 Myr ago and the younger one between 7 and 6 Myr ago. 60% of all faunal changes (immigrations/extinctions) that happened between 12 and 5.5 Myr ago occurred in these times of faunal turnover, when the rate of extinction/immigration increased up to four times the 'normal' value. With the exception of the extinction pattern of Siwalik rodents in the late Miocene epoch, there is a high correlation between the faunal turnover maxima of both the Spanish and the Siwalik herbivore communities, despite the transition from C₃ to C₄ plants at 8 Myr ago in Pakistan.

Cerling *et al.* reply — Köhler *et al.* suggest that phenomena other than floral change may be involved in the late Miocene global vegetation change, such as monsoonal dynamics or unnamed “other factors”. Citing evidence from Spain and Pakistan, they do not believe that there is necessarily a synchronicity or a causal link between faunal and vegetation change in the late Miocene epoch. However, on the contrary, it seems highly unlikely that a vegetation change on the scale documented¹ would be uncorrelated with faunal change.

Widespread faunal change in the late Miocene epoch was recognized^{7,10–12} long before the carbon-isotope shift was identified; our work was the first to attempt to link these widespread faunal changes to global vegetation change^{1,13}. For example, in North America, Webb *et al.*¹⁴ state that the “boundary between the Early and Late Hemphillian (about 6 Myr ago) records a mass extinction event for equids, when about ten of the existing 18 lineages vanished”. Although it is difficult to ‘prove’ causality in historical events, it seems likely that widespread faunal changes are linked to widespread vegetation changes.

The data from the Siwalik sediments in Pakistan are especially informative, because only from this region are there coeval data on faunal turnover, isotope palaeoecology, and upwelling related to monsoon dynamics (Fig. 1). Smoothed palaeosol data for carbon-13 content ($\delta^{13}\text{C}$) show a sharp change starting about 7 Myr ago and continuing to about 5 Myr ago, denoting the shift from C_3 - to C_4 -dominated vegetation.

The $\delta^{13}\text{C}$ data for tooth enamel show that the dietary change, which enhances the C_3 or C_4 signal by selective feeding, can be seen somewhat earlier than in the palaeosols, a result to be expected. Smoothed $\delta^{18}\text{O}$ data from palaeosols indicates a change in soil waters that precedes the $\delta^{13}\text{C}$ shift and which is correlated with increased abundance of upwelling indicators in the Arabian Sea at about 8.5 Myr ago.

Therefore the isotope record in the Siwaliks records two signals: a change in monsoonal dynamics at about 8.5 Myr ago and a pronounced vegetation change at about 7 Myr ago. Detailed faunal collections from the same region document several important turnover events. The two biggest events are at about 7 and 8.5 Myr ago (Fig. 1) and correspond to the two periods of change recorded in the isotope record.

Although the record is indeed complicated, the stable isotope record documents two important events affecting faunal change in the Siwaliks: one starting about 8.5 Myr ago that is related to the monsoon intensification, and a slightly later event related to expansion in C_4 biomass. Earlier faunal changes, such as those before 10 Myr ago as mentioned by Köhler *et al.*, are unre-

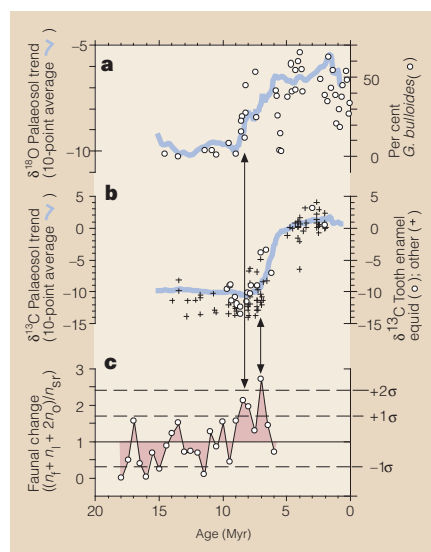


Figure 1 Data from Pakistan's Siwalik sediments show the two biggest events occurring at about 7 and 8.5 Myr ago. **a**, $\delta^{18}\text{O}$ data from Siwalik palaeosols, representing a trend determined by taking a 10-point running average of the roughly 200 palaeosols from the interval 16 to 0 Myr ago¹⁶. Also shown is the fraction of *Globigerina bulloides* from the Arabian Sea, an indicator of upwelling related to monsoon dynamics¹⁷. **b**, $\delta^{13}\text{C}$ data for palaeosols and for mammals' tooth enamel^{118–20} in the Siwaliks, representing a trend determined by taking a 10-point running average of the 200 or so palaeosols from the interval 16 to 0 Myr ago¹⁶. **c**, Faunal change index from the Siwaliks, represented by the number of first (n_1) and last (n_l) occurrences, including only occurrences (n_o), normalized to species richness (n_{sr}). Data from ref. 7. The index is normalized to 1.0 for the total data set.

lated to the global expansion of C_4 biomass.

C_4 photosynthesis is an adaptation to low atmospheric CO_2 levels. Because CO_2 gain and water loss both occur through stomata in C_3 plants, we expect that C_3 plants adapted to aridity would prosper in periods of lower atmospheric CO_2 . We would therefore expect that global changes within C_3 flora accompanied the C_4 expansion at the end of the Miocene epoch. Changes within C_3 ecosystems can be related to changes in atmospheric CO_2 levels (for example, the Pleistocene/Holocene transition¹⁵).

So, although C_4 plants did not flourish in Europe or in other high-latitude regions, it is likely that floral change occurred in those regions within C_3 ecosystems through the Miocene/Pliocene transition. The absence of evidence for C_4 expansion in Europe should not be taken to mean that floral change did not take place in Europe at the end of the Miocene; the isotope record is silent on that issue.

Thure E. Cerling

Department of Geology, University of Utah,
Salt Lake City, Utah 84112, USA

John M. Harris

Page Museum, Los Angeles, California 90036, USA

Bruce J. MacFadden

Florida Museum of Natural History, Gainesville,
Florida 32611, USA

Jay Quade

University of Arizona, Tucson, Arizona 85721, USA

Meave G. Leakey

National Museums of Kenya, Nairobi, Kenya

Vera Eisenmann

Musé National d'Histoire Naturelle,
75005 Paris, France

James R. Ehrlinger

Department of Biology, University of Utah,
Salt Lake City, Utah 84112, USA

1. Cerling, T. E. *et al.* *Nature* **389**, 153–158 (1997).
2. Agustí, J. *et al.* *J. Hum. Evol.* **31**, 143–155 (1996).
3. Garcés, M., Cabrera, L., Agustí, J. & Parés, J. M. *Geology* **25**, 19–22 (1997).
4. Garcés, M., Agustí, J., Cabrera, L. & Parés, J. M. *Earth Planet. Sci. Lett.* **142**, 381–396 (1996).
5. Moyà Solà, S. & Agustí, J. in *European Neogene Mammal Chronology* (eds Lindsay, E. H. *et al.*) 357–373 (Plenum, New York, 1990).
6. Johnson, N. M., Opdyke, N. D., Johnson, E. D., Lindsay, E. H. & Tahirkheli, R. A. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **37**, 17–44 (1982).
7. Barry, J. *et al.* *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **115**, 209–226 (1995).
8. Barry, J. in *Paleoclimate and Evolution* (eds Vrba, E., Denton, G. H., Partridge, T. C. & Burckler, L. H.) 115–134 (Yale Univ. Press, New Haven, 1995).
9. Cande, S. C. & Kent, D. V. *J. Geophys. Res.* **97**, 13917–13951 (1992).
10. Webb, S. D. *Ann. Rev. Ecol. Syst.* **9**, 393–426 (1977).
11. Janis, C. M. *Ann. Rev. Ecol. Syst.* **24**, 467–500 (1993).
12. Barry, J. *et al.* *Geology* **13**, 637–640 (1985).
13. Cerling, T. E., Ehrlinger, J. R. & Harris, J. M. *Proc. R. Soc. Lond.* (in the press).
14. Webb, S. D., Hulbert, R. C. & Lambert, W. D. in *Paleoclimate and Evolution, with Emphasis on Human Origins* (eds Vrba, E. *et al.*) (Yale Univ. Press, New Haven, 1995).
15. Jolly, D. & Haxeltine, A. *Science* **276**, 786–788 (1997).
16. Quade, J. & Cerling, T. E. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **115**, 91–116 (1995).
17. Kroon, D., Steens, T. & Troelstra, S. R. *Proc. ODP Sci. Res.* **117**, 257–263 (1991).
18. Cerling, T. E., Wang, Y. & Quade, J. *Nature* **361**, 344–345 (1993).
19. Morgan, M. E., Kingston, J. D. & Marino, B. D. *Nature* **367**, 162–165 (1994).
20. Stern, L. A., Johnson, G. D. & Chamberlain, C. P. *Geology* **22**, 419–422 (1994).

Life-support system benefits from noise

Mechanical ventilators are used to provide life support for patients with respiratory failure. But over the long term, these machines can damage the lungs, causing them to collapse and the partial pressure of oxygen in the arteries to drop to abnormally low values¹. In conventional mechanical ventilation, the respiratory rate and volume of air inspired per breath are fixed, although during natural breathing these parameters vary appreciably². A computer-controlled ventilator has now been introduced³ that can use noise to mimic this variability. We describe a conceptual model of lung injury in which the partial pressure of arterial oxygen is improved significantly by computer-controlled rather than conventional mechanical ventilation, in agreement with recent experimental data³.

To explain how variability can improve the arterial partial pressure of oxygen (pO_2), consider the pressure–volume (P – V) behaviour of an injured lung that is being mechanically ventilated with many peripheral airways closed, thereby creating large collapsed regions. Let α represent a fraction of the lung that is collapsed at the end of expiration. An uncollapsed lung will be ventilated according to a ‘normal’ nonlinear P – V relation⁴ (see the normalized P – V curve in Fig. 1a, labelled $\alpha = 0$). Collapsed regions, however, significantly alter the P – V curve⁵.

The limiting case of $\alpha = 1$ in Fig. 1a shows a model P – V curve for the first inflation of a completely collapsed lung, where V is proportional to P^N (N ranges from 10 to 16)⁶. When α is between 0 and 1, the P – V curve of the entire lung ($\alpha = 0.3$) will be a combination of the ‘normal’ curve and the P^N curve. Thus, for P values below 0.75, the highly nonlinear P^N term dominates, whereas, for P values above 0.75, the contribution of the ‘normal’ P – V curve leads to flattening of the P – V relation.

In conventional mechanical ventilation, P increases from end-expiratory pressure $P_{\text{exp}} = P_1$ (say $P_1 = 0.3$) to a fixed end-inspiratory pressure $P_{\text{ins}} = P_2$ (say $P_2 = 0.7$). The corresponding opened volume in the collapsed region increases from V_1 to V_2 . We mimic variability in breathing by adding noise to P_2 so that P increases from P_1 to $P_{\text{ins}} = P_2 + \eta$, where η is a random variable changing from breath to breath and is taken from a zero-mean gaussian distribution (Fig. 1b).

Suppose that, for one inflation, P increases to $P_{\text{ins}} = 0.75$ rather than to 0.7. This results in gaining recruited volume compared with $P_{\text{ins}} = 0.7$. Suppose now that for the next inflation, P increases to only $P_{\text{ins}} = 0.65$, losing some recruited volume. Owing to the strong nonlinearity (P^N) of the P – V curve, the ‘gain’ of volume for $P_{\text{ins}} > P_2$ is far greater than the ‘loss’ of volume for $P_{\text{ins}} < P_2$. When P_{ins} samples the gaussian around P_2 many times, the mean of P_{ins} will be P_2 , but the mean of the distribution of the recruited volumes will increase from V_2 to V_3 . The quantity $\Delta V = V_3 - V_2$ represents the net improvement, which is more than 240%.

Hence surface area for gas exchange in the collapsed region increases, leading to an increase in arterial pO_2 . In addition, as lung injury progresses, α increases, and the P – V curve of the entire lung gradually shifts towards the P^N limit. Therefore, with increasing α , adding noise to ventilation should increasingly improve the arterial pO_2 , a prediction that is consistent with experiments³.

The process of varying P around P_2 is analogous to the noise-enhanced amplification of a useful signal in a system by stochastic resonance⁷. In stochastic resonance, increasing the standard deviation (s.d.) of the noise in a nonlinear system will initially amplify a weak input so as to increase the

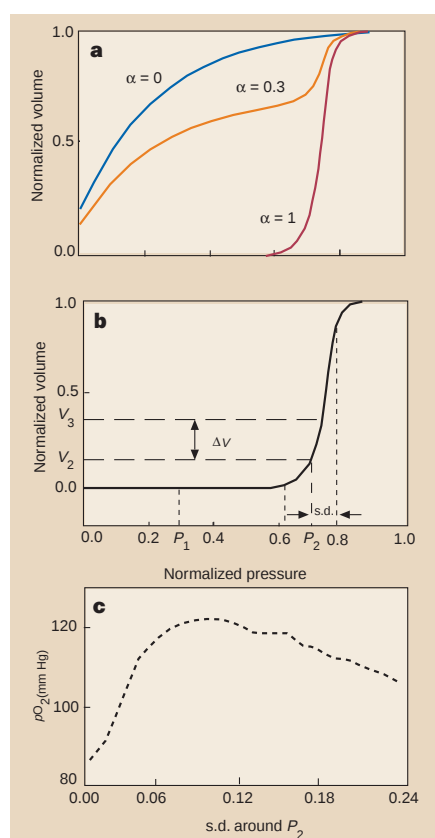


Figure 1 Variability improves arterial partial pressure of lung oxygen. **a**, Pressure–volume (P – V) curves normalized to unity at total lung capacity. $\alpha = 0$, normal P – V of a lung without collapsed regions⁴. $\alpha = 1$, P – V for a collapsed lung⁶ where recruitment of volume is proportional to P^N for $P < 0.75$. $\alpha = 0.3$, weighted average of the two limiting cases. **b**, Normalized P – V curve of a collapsed region (case $\alpha = 1$ from **a**). P_1 , end-expiratory pressure; P_2 , end-inspiratory pressure; V_2 , corresponding recruited volume. When noise (s.d.=0.075) is added to P_2 , average opened volume increases from $V_2 = 0.15$ to $V_3 = 0.363$. **c**, Predicted arterial blood oxygen partial pressure pO_2 as a function of the s.d. of the gaussian around $P_2 = 0.7$. pO_2 data obtained by calculating and averaging 1,000 normalized compliance values, C , which, using ref. 3 data, we relate to pO_2 ($pO_2 = 2.8C + 6$).

output signal-to-noise ratio; however, further increasing the standard deviation will have the opposite effect. The output signal in our case is the arterial pO_2 . When small noise is added to P_2 , the surface area for gas exchange, and hence arterial pO_2 , increases.

Increasing the noise amplitude too much may adversely affect the arterial pO_2 . For example, as we gradually increase the standard deviation of the gaussian noise along the S-shaped nonlinearity curve ($\alpha = 0.3$ in Fig. 1a), we find that the normalized compliance, C (defined as $V_T/(P_{\text{ins}} - P_1)$, where $P_{\text{ins}} = P_2 + \eta$, and V_T is the volume inspired per breath (corresponding to $P_{\text{ins}} - P_1$), displays a maximum.

As C is linearly related to arterial pO_2 in lung injury³ (probably because the collapse of lung regions leads to proportional changes

in the area available for gas exchange), our model predicts that there is an optimum standard deviation at which pO_2 also displays a maximum (Fig. 1c). So the possibility of tuning noise for optimal gas exchange in mechanical ventilation arises, from the presence of a nonlinearity due to the competing effects of recruitment of alveoli via avalanches⁸ (causing C to increase) and the gradual stiffening of the overinflated parenchymal tissues⁴ (causing C to decrease).

As well as offering immediate improvement in gas exchange, noise may have long-term benefits for patients with acute lung injury and respiratory failure because, without requiring increased mean airway pressures, fewer alveolar regions will remain collapsed. This is significant, as high airway pressures cause mechanical failure of pulmonary microvasculature⁹, and high shear forces on the alveolar walls increase the level of inflammation which can further propagate the inflammatory response within the alveolar compartment¹⁰. So including appropriately designed noise in mechanical ventilators will improve gas exchange and could have a significant effect on morbidity by breaking the chain of injury propagation in acute lung injury.

B. Suki, A. M. Alencar, M. K. Sujeer, K. R. Lutchen, J. J. Collins

Department of Biomedical Engineering, Boston University, Boston, Massachusetts 02215, USA
e-mail: bs@enga.bu.edu

J. S. Andrade, Jr

Departamento de Física, Universidade Federal do Ceará, 60451-970 Fortaleza, Ceará, Brazil

E. P. Ingenito

Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA

S. Zapperi, H. E. Stanley

Department of Physics, Boston University, Boston, Massachusetts 02215, USA

1. Dreyfuss, D., Soler, P., Basset, G. & Saumon, G. *Am. Rev. Resp. Dis.* **137**, 1159–1164 (1988).
2. Dejours, P., Puccinelli, R., Armand, J. & Dicharry, M. *Respir. Physiol.* **1**, 265–280 (1966).
3. Lefevre, G. R., Kowalski, S. E., Girling, L. G., Thiessen, D. B. & Mutch, W. A. C. *Am. J. Respir. Crit. Care Med.* **154**, 1567–1572 (1996).
4. Salazar, E. & Knowles, J. H. *J. Appl. Physiol.* **19**, 97–104 (1964).
5. Cheng, W., DeLong, D. S., Franz, G. N., Petsonck, E. L. & Frazer, D. G. *Respir. Physiol.* **102**, 205–215 (1995).
6. Sujeer, M. K. et al. *Phys. Rev. E* **56**, 3385–3394 (1997).
7. Wiesenfeld, K. & Moss, F. *Nature* **373**, 33–36 (1995).
8. Suki, B., Barabási, A. L., Hantos, Z., Peták, F. & Stanley, H. E. *Nature* **368**, 615–618 (1994).
9. Costello, M. L., Mathieu-Costello, O. & West, J. B. *Am. Rev. Resp. Dis.* **145**, 1446–1455 (1992).
10. Tremblay, L., Valenza, F., Ribeiro, S. P., Li, J. & Slutsky, A. S. *J. Clin. Invest.* **99**, 944–952 (1997).

correction

In ‘What’s so special about figs?’ (*Nature* **392**, 668; 1998) the values given in Table 1 for copper, iron, manganese and zinc should have been expressed as μg per g dry matter. Also, in ref. 1, the first author’s name should read ‘Conklin, N. L.’

Raising a glass to ale

Beer: Tap into the Art and Science of Brewing

by Charles Bamforth

Plenum: 1998. Pp. 245. \$27.95, £18.50

John Postgate

Wild yeasts are abundant in nature, finding the sugary substrates they favour in fruit juices, the sap of many plants, nectar and honey. It is not difficult to guess how people acquired, early in human history, a taste for such fermented liquors as wines, cider and mead: their crude equivalents would have appeared spontaneously, ready for the potential brewster to discover, imitate and adapt to the liking of her family group. But how anyone came to develop (over 8,000 years ago, it seems) as complicated a beverage as beer, which is the fermented extract of a partly germinated grain, taxes even a Darwinian imagination, accustomed to contemplating the evolution of complexity.

Charles Bamforth has a stab at an explanation — that beer arose as an offshoot of baking technology — but the bulk of this entertaining and informative book is concerned less with historical matters than with the aforesaid complexity: the production of beer today. In outline, beer is made by allowing barley to germinate, so that hydrolysis of its storage polysaccharide, starch, begins. Germination is then halted by drying, and the nascent seedling is milled and extracted with hot water. This extract, rich in sugars, dextrans and residual biochemicals, is fermented with carefully preserved strains of yeast. Various filtration, heating, chilling, sedimentation, clarification and maturation steps are involved, and flavourings, notably hops, may be added. The ultimate product is then barrelled, bottled, canned or otherwise packaged for marketing.

Bamforth, who is deputy director of the United Kingdom's Brewing Research International (the erstwhile Brewing Industry Research Foundation at Nutfield in Surrey), takes one gently through every facet of contemporary brewing in what is a triumph of user-friendly exposition: he writes in an unpretentious but authoritative mid-Atlantic style; he assumes very little background knowledge; and if too many details seem to be piling up, he has some anecdote or aside with which to recapture one's attention. He comes across as an amiable guide conducting one around a site he knows well and loves. Some scientific and technical terminology is inescapable, but Bamforth is careful to explain such terms at first mention, and provides a comprehensive glossary at the end. There is also an appendix, which amplifies a few scientific concepts that would have been too much of a digression in the main text.



ANTHONY BLAKE PHOTO LIBRARY

Thirsty? The UK's taste for real ale has refreshed the brewing industry.

Apart from his central theme of mainstream brewing, Bamforth deals with all sorts of related topics. For example, he explains how a widgeot makes beer from a can taste like proper draught (I think it does, by the way; I dutifully tested an example); that brown bottles preserve beer better than green or clear ones (they inhibit a deleterious photochemical change); that a few centuries ago, hops were regarded as wicked and pernicious adulterants of good ale; that Budweiser includes rice in its malt; that the famous catchphrase "Guinness is good for you" is no longer legal; and that both ancient Egyptian and genetically engineered beers have been brewed recently.

Bamforth maintains that brewing has become an interdisciplinary science, drawing from chemistry, engineering and microbiology. I cannot for a moment agree with this proposition, even in the light of his masterly account; to me, brewing remains manifestly an art. Certainly an enormous amount of science has stemmed from it; the contribution brewing research has made to analytical chemistry, biochemistry and microbiology is inestimable, and much basic science of universal importance has emerged from such laboratories as Nutfield or the Carlsberg Institute. Science, too, has had an enormous

impact on the technology the modern brewer or maltster can command. But it is the experience, intuition and judgement of the brewer that counts.

Brewing beer for a mass market has become so sophisticated that in the United Kingdom — and this reflects a worldwide trend — the 6,477 independent breweries that existed in 1900 had dwindled by 1980 to 191. Since then an emergence of small, specialist breweries brought the number up to 499 by 1996, helped no doubt by an expanding market for 'real ale' (not that the mass-produced beverage is in any sense unreal!).

Bamforth's book is an ideal introductory work for anyone entering the industry, whether in brewing itself, in brewing research, as a publican, in marketing, or in supply areas such as hop or barley farming. It is also excellent for students of applied microbiology and bio-engineering. But its easy style ought also to attract a wider audience: technical and scientific journalists will find it immensely helpful, and it would make enjoyable holiday reading for microbiologists, chemists, engineers and, of course, the average beer drinker. □

John Postgate is emeritus professor of Microbiology at the University of Sussex, Brighton BN1 9RH, UK.

Concise cosmology

The Little Book of the Big Bang: A Cosmic Primer

by Craig Hogan

Copernicus: 1998. Pp. 181. \$20, £15.50

George Ellis

The explosion of data and theory in cosmology has led to a profusion of books about modern cosmology written for non-experts. Craig Hogan's cosmic primer ranks among the best: it is a gem. It is concisely written and very carefully crafted, with useful summary tables and diagrams. From a visual viewpoint it is restrained (there are no colour illustrations or photographs), but it provides a superb overview of cosmic processes (for example, nucleosynthesis and the origin of the cosmic background radiation spectrum). There are many interesting insights deriving from Hogan's deep involvement with the field; these will interest professionals as well as newcomers to cosmology.

Hogan raises and answers the main questions the uninitiated want answered: "Is everything expanding?", "What is the Universe expanding into?", "What happened before the Big Bang?", and so forth. He draws many excellent informative analogies, for example pointing out that the dynamical

equations of the expanding Universe are the same as those of a ball thrown vertically from the surface of the Earth, so the same set of possibilities arise in its solutions.

But there are weak points. An issue regarding the book's presentation is related to a fundamental dilemma in any course or text: whether to present the material in the order that is easiest to understand or in a more logical order that is more satisfying but harder to get to grips with. Hogan has chosen the logical approach. As a result, many readers who would appreciate the easily grasped later sections on the Big Bang—which are accessible enough for high-school students and the general public—may well fall by the wayside during the second and third chapters; these contain an excellent but technically demanding summary of space and time scales and modern fundamental physics, for which a first-year university physics course is probably required. Readers who struggle with these should skip them at first reading and return to them later; ideally these chapters should be shifted to the end of the book.

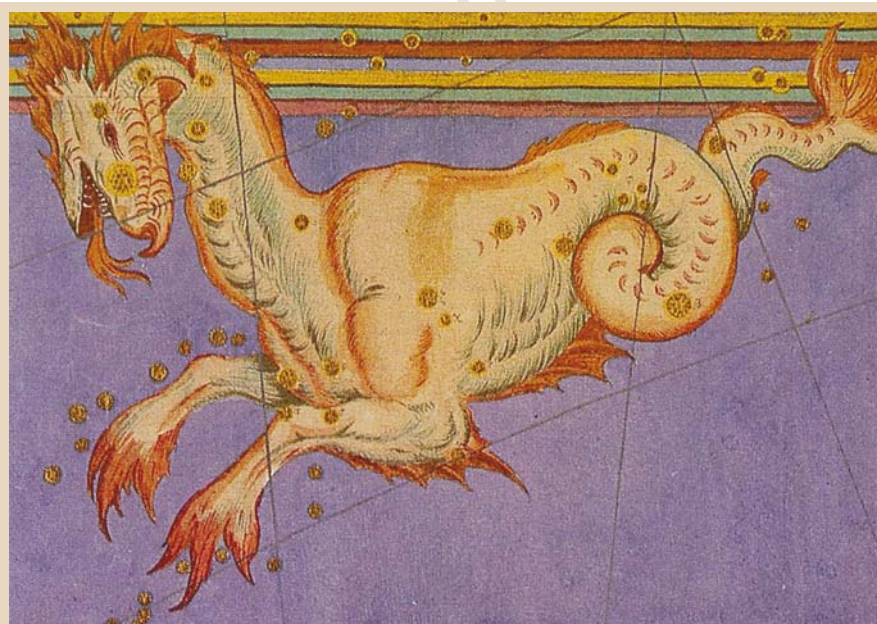
Of broader concern is the degree of certainty with which various theories are presented. Hogan is generally careful about this, for example emphasizing that we cannot obtain evidence about events outside our past light cone (so discussions of the nature

of the Universe on a super-Hubble scale are of a rather academic nature), but he restricts his discussion to a critical-density inflationary Universe, and in various places presents the inflationary Universe idea as a well-established theory. As he says himself, the relevant physics at times before the electro-weak epoch is not well established; inflation is still a hypothesis rather than a fact. Furthermore, the density of matter in the universe is probably less than the critical value. A few more cautionary notes about this proposal would be in order.

A brief section at the end relates to the meaning of the Universe and anthropic issues. Although the technical side of this discussion is well done, any treatment of ideas of self and free will (as briefly mentioned here) requires a more extended and philosophically developed discussion. For example, Hogan states that "the most important thing about us is what we create: our perceptions and technologies, our culture and societies, our art and our science." This is a highly debatable philosophical position; many would claim in contrast that the most important thing about us is how we act, expressing our ethical stance and moral purpose. In discussing meaning-of-life issues, one cannot avoid such terrain.

There is a misleading diagram that represents our past light cone as like a conical hat, the apex being the present day, flaring out at the bottom towards the surface of last scattering. The ever-increasing size of the light cone as one goes back into the past is because the diagram represents the present-day distances of the objects concerned. But this is completely different from their distances when they emitted the light, so the actual shape of the past light cone is quite different from that shown. Its true shape is like an onion with the sharp edge upwards; it flares out to a maximum radius (at a redshift of 1.25 in a critical-density Universe) and then plunges sharply back in towards the hot early phase. In my view it would be much better to have that shape represented in this diagram. The figure does not show the maximum size of the past light cone in the past, nor the associated maxima of angular separation of null geodesics that result in angular sizes of rigid rods having a minimum.

A second technical issue is the claim that, in a Universe that lasts forever, despite infinite time being available, there would not be enough time to think all possible thoughts. This conclusion is based on allowing the coding of these thoughts to require an infinite chain of symbols, but that does not make sense. It must be possible to code any thought that is meaningful in a finite number of symbols, for otherwise it cannot be processed (we can think about infinity, for example, precisely because it can be coded in a single symbol). There are effectively



Publishing constellations

The constellation Cetus, from the sky atlas by Bayer (1603), as reproduced in Jay M. Pasachoff's *The Peterson Field Guide to the Stars and Planets* (Houghton Mifflin, \$18 (pbk)). It joins several other recent astronomical and cosmological primers, including Arthur Upgren's *Night Has a Thousand Eyes: A Naked-Eye Guide to the Sky, Its Science, and Lore*

(Plenum, \$27.95), T. Padmanabhan's *After the First Three Minutes: The Story of Our Universe* (CUP, £35, \$59.95 (hbk); £12.95, \$19.95 (pbk)), Marcelo Gleiser's *The Dancing Universe: From Creation Myths to the Big Bang* (Dutton, \$25.95) and Timothy Ferris's *The Whole Shebang: A State-of-the-Universe(s) Report* (Weidenfeld/Simon and Schuster, £20, \$25).

only a (very large) finite number of possible distinct brain states in any brain of finite size.

These comments do not detract from the overall value of the presentation, which I warmly recommend. I am sure there will be many further editions of this fine book. □

George Ellis is in the Mathematics Department, University of Cape Town, Rondebosch 7700, Cape Town, South Africa.

Sex and sensibility

The Two Sexes: Growing Up Apart, Coming Together

by Eleanor E. Maccoby

Harvard University Press: 1998. Pp. 376.

\$39.95, £26.50

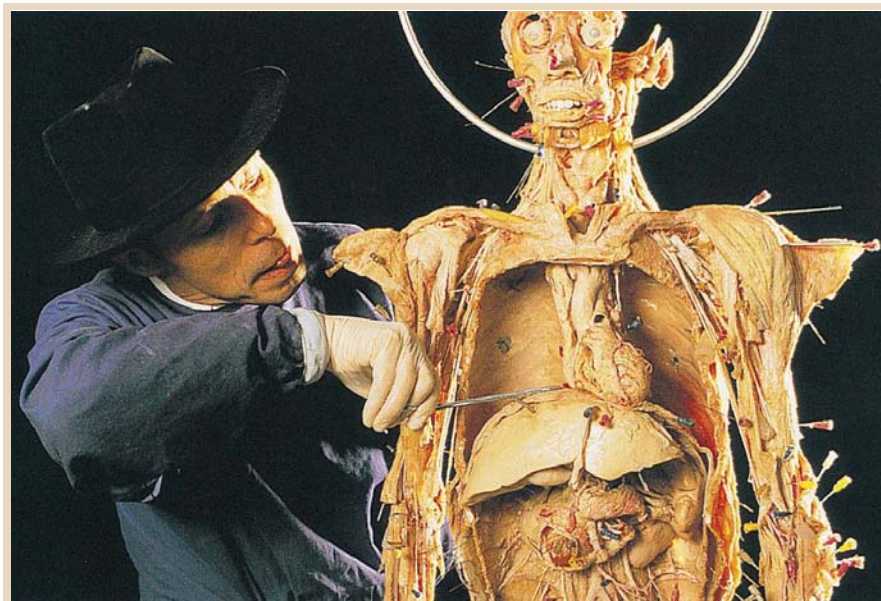
David H. Skuse

Are men really from Mars and women from Venus? In *The Two Sexes*, Eleanor E. Maccoby attempts to explain the considerable evidence that male and female patterns of social interaction differ from one another in terms of both same-sex and opposite-sex exchanges.

In 1974, Maccoby published with Carol Jacklin a landmark work on the psychology of sex differences, which arguably heralded the birth of a movement in feminist psychology. Stimulated by the conclusions of that synthesis of previous research, which claimed such differences were smaller than had been generally assumed, feminist scholars set out to show that, if gender studies were conducted with due regard to cultural relativity and gender stereotyping (by subject and observer), consistent findings would emerge. They reasoned that the results would show few, small or no differences in cognitive and social characteristics by sex.

Many meta-analyses later, only the most diehard psychologist would still cling to the view that sex differences in behaviour do not exist. Meanwhile, Maccoby has become a proponent of the popular 'separate cultures' school, which argues that children learn rules for social interaction from experience in largely peer-aggregated groups. They carry this learning through to adulthood. So when a man's wife asks him "What's wrong, honey?", and he replies "Nothing", and she persists "I can tell you're annoyed with me — I can see it in your face", he retorts "I tell you, there's nothing wrong, cut it out won't you!" because he is male, and that is how he has learned to respond to questions about his emotional state.

Maccoby's book is beautifully written and organized. In the first half she takes us on a journey from infancy through pre-school behaviour, to elementary school and adolescence. Fascinating evidence is added for the formation of sex-segregated social groups, which are amazingly resistant



Insides out

The German anatomist Gunther von Hagens preserves human bodies by a process he calls 'plastination', in which he replaces all of the water and fat with silicone and other polymers. He began by creating anatomical preparations for educational purposes, but now also exhibits them at controversial shows. The corpses are obtained through a donation programme.

This image by Marc Steinmetz of von Hagens at work is from the science and technology section of 1998 *World Press Photo* (Thames and Hudson, £12.95 (pbk)), a remarkable collection of the previous year's most striking examples of photojournalism. The book brings together the winners of the 41st World Press Photo Contest.

in middle childhood to well-meaning adult attempts to force integration. In the second half, which is rather weaker in terms of argument and grasp of detail, issues of sex roles in the workplace and in parenthood are discussed. A variety of explanations for the seemingly universal existence of sex-role differentiation are proposed, from biology to evolutionary psychology, but Maccoby is most comfortable and knowledgeable when discussing the role of social influences.

As writers on the subject acknowledge, despite a tremendous and growing literature there has been no really satisfactory theoretical framework to encompass both the intrinsic differences in the ways the sexes behave and the environmental influences that moderate or mediate such traits. Although evolutionary psychology claims the high ground here, many commentators feel that retrospective predictions fail to do justice to the complexity of the issues.

Take the well-documented male propensity to dominate not only his own sex but particular females too. How could this be linked to other typically male characteristics, such as a lack of emotional expression (relative to females) and the irritating habit of interrupting people (up to five times as often as women in small-group interactions)? Do these features of 'male' behaviour arise innately, from 'essential qualities',

or are they by-products of socialization experiences?

Maccoby is rather surer about the origins of male behaviour than female behaviour. Gender divergence, she argues, has its origins in the way children of a given sex aggregate with children of the same age. Groups of male children tend to be more cohesive than female aggregations, and their play style (which she characterizes as rough and essentially physical in nature) is highly attractive to boys but is a turn-off to most girls. On the other hand, female groups are loose-knit affairs, although they are also formed on the basis of play-style compatibility and are, as any male will tell you, not at all attractive to boys. It is implicit that girls form groups with girls largely because they do not, as a rule, want to join in with the boys.

Could it be that boys are primed by androgen exposure *in utero* to behave in a way that fosters group cohesion? Maccoby seems happy to posit biological factors as the ultimate explanation for male social behaviour, but for girls an environmental explanation is proposed. She thinks that girls are influenced by their upbringing to be more sensitive to others' feelings and to be more nurturing, socially sensitive, friendly and concerned with another's welfare.

Maccoby believes the differentiation of social behaviour by sex in adulthood

reflects 'echoes of childhood'. Perhaps we need to repress styles of social interaction learned then in order to function adequately as parents, as she suggests. But we are still some way from a comprehensive explanation, as opposed to a mere description, of the way the sexes differ in their behaviour. The popularity of books lamenting the difficulty men and women have in understanding each other's verbal and non-verbal discourse is testimony to that sad fact. □

David H. Skuse is in the Behavioural Sciences Unit, Institute of Child Health, 30 Guilford Street, London WC1N 1EH, UK.

Students' dilemma

Essential Cell Biology: An Introduction to the Molecular Biology of the Cell

by Bruce Alberts, Dennis Bray, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts and Peter Walter

Garland: 1998. Pp. 630. \$59.95, £24.95

David S. Latchman

Even before I opened this book, the fact that it was from the authors of the highly acclaimed *Molecular Biology of the Cell* led me to expect it to be outstanding in both the lucidity of the text and the clarity of the illustrations. I was not disappointed. If anything, this book is even easier to follow than its predecessor.

A conscious effort to make things as simple as possible seems to have been made: the subheadings for each section make the appropriate point simply and effectively, and the order of the chapters has been carefully thought out. So, unlike *Molecular Biology of the Cell*, the chapter on mitosis and meiosis precedes that on the cell cycle, which is, in my opinion, much better for undergraduates. The work is generally up to date, although if it was written today I suspect there would be at least a mention of the effect of histone acetylation on gene expression, and the cloning of Dolly would receive more than a brief mention in a figure legend.

But overall this is an excellent work, and normally I would recommend it wholeheartedly. Given the success of *Molecular Biology of the Cell*, however, one must ask how it is intended to be used in relation to the previous book. The introduction to *Essential Cell Biology* says it should be "easily understood by first or second year undergraduates with little background in biology". Apparently *Molecular Biology of the Cell* is aimed at "advanced undergraduates specialising in the life sciences or medicine". It was not always so: the preface to the first edition, which is still included in the current edition, says it was intended "for students taking a first course in cell biology" and that "even a stranger to biology could follow it by starting

at the beginning". Perhaps the authors changed their opinion as the field advanced and they received comments from readers.

Nonetheless, it would have been helpful to have been told the target audience for each of these texts. A student taking an isolated introductory course would buy *Essential Cell Biology* and find it highly appropriate. But should a student taking an introductory course who wants to go further and take specialist courses buy *Essential Cell Biology* and then *Molecular Biology of the Cell*, even though there is considerable duplication, or risk buying just the more complex *Molecular Biology of the Cell*?

The excellence of both books makes this question all the more pressing. The clarity and attractiveness of *Essential Cell Biology* will surely ensure it a worldwide audience in the same way as *Molecular Biology of the Cell*. But the financially and academically hard-pressed student is entitled to know which book they should purchase and when. □

David S. Latchman is in the Windeyer Institute of Medical Sciences, University College London Medical School, Windeyer Building, 46 Cleveland Street, London W1P 6DB, UK.

At a glance

Nucleosynthesis and Chemical Evolution of Galaxies

by Bernard E. J. Pagel

Cambridge University Press: 1997. Pp. 378. £19.95, \$29.95 (pbk); £55, \$74.95 (hbk)

Take 10^{11} solar masses of gas with almost primordial composition, add many generations of stars formed at a suitable rate, let the elements be synthesized inside the stars, add some heavy elements in ejecta from dying stars, and stir the ejecta with the surrounding medium over a low heat for 10^{10} years. The result should be the pattern of elemental abundances observed in our own Galaxy and external galaxies. Surprisingly, the cosmic soup tastes the same (with a few exceptions) almost everywhere. Why?

This subject is addressed by Bernard Pagel as he guides the reader magisterially through the secrets of measuring elemental abundances and producing the first light elements by cosmological nucleosynthesis, and later all the others by stellar processing (primary and secondary elements, hydrostatic and explosive phases, in single and binary stars and through neutron-capture synthesis). He also discusses the concepts of galactic chemical evolution and its theoretical modelling in steps of different complexity, and finally the comparison of the observational data with the theoretical results for galaxies of different morphological type and star-formation history. Each topic is explained from the ground up, giving the reader all the information needed to proceed further. The book manages to present (without losing scientific rigour) one of the most fascinating subjects of modern astrophysics. □

Cesare Chiosi

Dipartimento di Astronomia, Vicolo Osservatorio 5, 35122 Padova, Italy.

New Journals

This year, *Nature's* annual new journals review supplement will appear in the issue of 10 September. Publishers and learned societies are invited to submit journals for review, as well as details of any eligible electronic journals, taking note of the following criteria:

- Journals must have first appeared during or after June 1996 and issued at least four separate numbers by the end of May 1998.
- Journals covering any aspect of science are eligible, although those dealing with clinical medicine and pure mathematics are excluded, as are newsletters and publications of abstracts.
- Frequency of publication must be at least three times a year.
- The main language is English.
- Deadline for submission is 5 June.

Please send at least four different issues (the first, the most recent and any two others) of each eligible title, together with full details of subscription rates, to: Peter Tallack, *Nature*, Porters South, Crinan Street, London N1 9XW, UK. Tel: +44 (0)171 843 4567. e-mail: p.tallack@nature.com

Mitosis and Apoptosis: Matters of Life and Death

by I. D. Bowen, S. M. Bowen and A. H. Jones
Chapman and Hall: 1997. Pp. 182. £24.99, \$49.95 (pbk)

A book of modest size but great ambition, this endeavours to explain cell proliferation and cell death at the molecular, cellular, physiological and philosophical levels. It is aimed at students at graduate and advanced undergraduate levels, and covers a broader area, but in less depth, than the title suggests.

For example, in addition to the mechanics of cell-cycle control and apoptosis, one of the six chapters is devoted to signal transduction. Because signal transduction controls both proliferation and death, this chapter nicely links the two topics. But there is a price to pay. In exchange for the broad coverage, the reader gets short-changed on several concepts central to cell division and cell death.

There is no serious discussion of chromosome dynamics, and DNA replication is covered in just a few brief paragraphs.

The chapters on cell death are more complete, including some useful discussions on non-apoptotic types of death. In general, the book is fairly up to date and contains many references to primary research articles. If a brief introduction to cell proliferation and cell death is required, this concise, enjoyable book might well be it. □

Michael O. Hengartner

Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA.
e-mail: mir@cshl.org

A silicon-based nuclear spin quantum computer

B. E. Kane

Semiconductor Nanofabrication Facility, School of Physics, University of New South Wales, Sydney 2052, Australia

Quantum computers promise to exceed the computational efficiency of ordinary classical machines because quantum algorithms allow the execution of certain tasks in fewer steps. But practical implementation of these machines poses a formidable challenge. Here I present a scheme for implementing a quantum-mechanical computer. Information is encoded onto the nuclear spins of donor atoms in doped silicon electronic devices. Logical operations on individual spins are performed using externally applied electric fields, and spin measurements are made using currents of spin-polarized electrons. The realization of such a computer is dependent on future refinements of conventional silicon electronics.

Although the concept of information underlying all modern computer technology is essentially classical, physicists know that nature obeys the laws of quantum mechanics. The idea of a quantum computer has been developed theoretically over several decades to elucidate fundamental questions concerning the capabilities and limitations of machines in which information is treated quantum mechanically^{1,2}. Specifically, in quantum computers the ones and zeros of classical digital computers are replaced by the quantum state of a two-level system (a qubit). Logical operations carried out on the qubits and their measurement to determine the result of the computation must obey quantum-mechanical laws. Quantum computation can in principle only occur in systems that are almost completely isolated from their environment and which consequently must dissipate no energy during the process of computation, conditions that are extraordinarily difficult to fulfil in practice.

Interest in quantum computation has increased dramatically in the past four years because of two important insights: first, quantum algorithms (most notably for prime factorization^{3,4} and for exhaustive search⁵) have been developed that outperform the best known algorithms doing the same tasks on a classical computer. These algorithms require that the internal state of the quantum computer be controlled with extraordinary precision, so that the coherent quantum state upon which the quantum algorithms rely is not destroyed. Because completely preventing decoherence (uncontrolled interaction of a quantum system with its surrounding environment) is impossible, the existence of quantum algorithms does not prove that they can ever be implemented in a real machine.

The second critical insight has been the discovery of quantum error-correcting codes that enable quantum computers to operate despite some degree of decoherence and which may make quantum computers experimentally realizable^{6,7}. The tasks that lie ahead to create an actual quantum computer are formidable: Preskill⁸ has estimated that a quantum computer operating on 10^6 qubits with a 10^{-6} probability of error in each operation would exceed the capabilities of contemporary conventional computers on the prime factorization problem. To make use of error-correcting codes, logical operations and measurement must be able to proceed in parallel on qubits throughout the computer.

The states of spin $1/2$ particles are two-level systems that can potentially be used for quantum computation. Nuclear spins have been incorporated into several quantum computer proposals^{9–12} because they are extremely well isolated from their environment and so operations on nuclear spin qubits could have low error rates. The primary challenge in using nuclear spins in quantum computers lies in measuring the spins. The bulk spin resonance approach

to quantum computation^{11,12} circumvents the single-spin detection problem essentially by performing quantum calculations in parallel in a large number of molecules and determining the result from macroscopic magnetization measurements. The measurable signal decreases with the number of qubits, however, and scaling this approach above about ten qubits will be technically demanding³⁷.

To attain the goal of a 10^6 qubit quantum computer, it has been suggested that a 'solid state' approach¹³ might eventually replicate the enormous success of modern electronics fabrication technology. An attractive alternative approach to nuclear spin quantum computation is to incorporate nuclear spins into an electronic device and to detect the spins and control their interactions electronically¹⁴. Electron and nuclear spins are coupled by the hyperfine interaction¹⁵. Under appropriate circumstances, polarization is transferred between the two spin systems and nuclear spin polarization is detectable by its effect on the electronic properties of a sample^{16,17}. Electronic devices for both generating and detecting nuclear spin polarization, implemented at low temperatures in GaAs/Al_xGa_{1-x}As heterostructures, have been developed¹⁸, and similar devices have been incorporated into nanostructures^{19,20}. Although the number of spins probed in the nanostructure experiments is still large ($\sim 10^{11}$; ref. 19), sensitivity will improve in optimized devices and in systems with larger hyperfine interactions.

Here I present a scheme for implementing a quantum computer on an array of nuclear spins located on donors in silicon, the semiconductor used in most conventional computer electronics. Logical operations and measurements can in principle be performed independently and in parallel on each spin in the array. I describe specific electronic devices for the manipulation and measurement of nuclear spins, fabrication of which will require significant advances in the rapidly moving field of nanotechnology. Although it is likely that scaling the devices proposed here into a computer of the size envisaged by Preskill⁸ will be an extraordinary challenge, a silicon-based quantum computer is in a unique position to benefit from the resources and ingenuity being directed towards making conventional electronics of ever smaller size and greater complexity.

Quantum computation with a ^{31}P array in silicon

The strength of the hyperfine interaction is proportional to the probability density of the electron wavefunction at the nucleus. In semiconductors, the electron wavefunction extends over large distances through the crystal lattice. Two nuclear spins can consequently interact with the same electron, leading to electron-mediated or indirect nuclear spin coupling¹⁵. Because the electron is sensitive to externally applied electric fields, the hyperfine inter-

action and electron-mediated nuclear spin interaction can be controlled by voltages applied to metallic gates in a semiconductor device, enabling the external manipulation of nuclear spin dynamics that is necessary for quantum computation.

The conditions required for electron-coupled nuclear spin computation and single nuclear spin detection can arise if the nuclear spin is located on a positively charged donor in a semiconductor host. The electron wavefunction is then concentrated at the donor nucleus (for s orbitals and energy bands composed primarily of them), yielding a large hyperfine interaction energy. For shallow-level donors, however, the electron wavefunction extends tens or hundreds of ångströms away from the donor nucleus, allowing electron-mediated nuclear spin coupling to occur over comparable distances. The quantum computer proposed here comprises an array of such donors positioned beneath the surface of a semiconductor host (Fig. 1). A quantum mechanical calculation proceeds by the precise control of three external parameters: (1) gates above the donors control the strength of the hyperfine interactions and hence the resonance frequency of the nuclear spins beneath them; (2) gates between the donors turn on and off electron-mediated coupling between the nuclear spins¹³; (3) a globally applied a.c. magnetic field B_{ac} flips nuclear spins at resonance. Custom adjustment of the coupling of each spin to its neighbours and to B_{ac} enables different operations to be performed on each of the spins simultaneously. Finally, measurements are performed by transferring nuclear spin polarization to the electrons and determining the electron spin state by its effect on the orbital wavefunction of the electrons, which can be probed using capacitance measurements between adjacent gates.

An important requirement for a quantum computer is to isolate the qubits from any degrees of freedom that may lead to decoherence. If the qubits are spins on a donor in a semiconductor, nuclear spins in the host are a large reservoir with which the donor spins can interact. Consequently, the host should contain only nuclei with spin $I = 0$. This simple requirement unfortunately eliminates all III–V semiconductors as host candidates, because none of their constituent elements possesses stable $I = 0$ isotopes²¹. Group IV semiconductors are composed primarily $I = 0$ isotopes and can in principle be purified to contain only $I = 0$ isotopes. Because of the

advanced state of Si materials technology and the tremendous effort currently underway in Si nanofabrication, Si is the obvious choice for the semiconductor host.

The only $I = 1/2$ shallow (group V) donor in Si is ^{31}P . The Si- ^{31}P system was exhaustively studied 40 years ago in the first electron–nuclear double-resonance experiments^{22,23}. At sufficiently low ^{31}P concentrations at temperature $T = 1.5$ K, the electron spin relaxation time is thousands of seconds and the ^{31}P nuclear spin relaxation time exceeds 10 hours. It is likely that at millikelvin temperatures the phonon limited ^{31}P relaxation time is of the order of 10^{18} seconds (ref. 24), making this system ideal for quantum computation.

The purpose of the electrons in the computer is to mediate nuclear spin interactions and to facilitate measurement of the nuclear spins. Irreversible interactions between electron and nuclear spins must not occur as the computation proceeds: the electrons must be in a non-degenerate ground state throughout the computation. At sufficiently low temperatures, electrons only occupy the lowest energy-bound state at the donor, whose twofold spin degeneracy is broken by an applied magnetic field B . (The valley degeneracy of the Si conduction band is broken in the vicinity of the donor²⁵. The lowest donor excited state is approximately 15 meV above the ground state²³.) The electrons will only occupy the lowest energy spin level when $2\mu_B B \gg kT$, where μ_B is the Bohr magneton. (In Si, the Landé g -factor is very close to +2, so $g = 2$ is used throughout this discussion.) The electrons will be completely spin-polarized ($n_\uparrow/n_\downarrow < 10^{-6}$) when $T \leq 100$ mK and $B \geq 2$ tesla. A quantum-mechanical computer is non-dissipative and can consequently operate at low temperatures. Dissipation will arise external to the computer from gate biasing and from eddy currents caused by B_{ac} , and during polarization and measurement of the nuclear spins. These effects will determine the minimum operable temperature of the computer. For this discussion, I will assume $T = 100$ mK and $B = 2$ T. Note that these conditions do not fully polarize the nuclear spins, which are instead aligned by interactions with the polarized electrons.

Magnitude of spin interactions in Si- ^{31}P

The size of the interactions between spins determines both the time

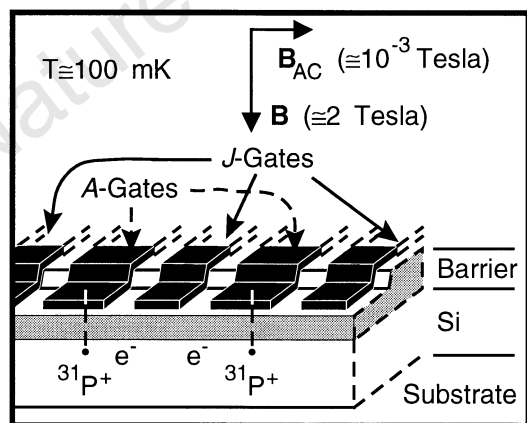


Figure 1 Illustration of two cells in a one-dimensional array containing ^{31}P donors and electrons in a Si host, separated by a barrier from metal gates on the surface. 'A gates' control the resonance frequency of the nuclear spin qubits; 'J gates' control the electron-mediated coupling between adjacent nuclear spins. The ledge over which the gates cross localizes the gate electric field in the vicinity of the donors.

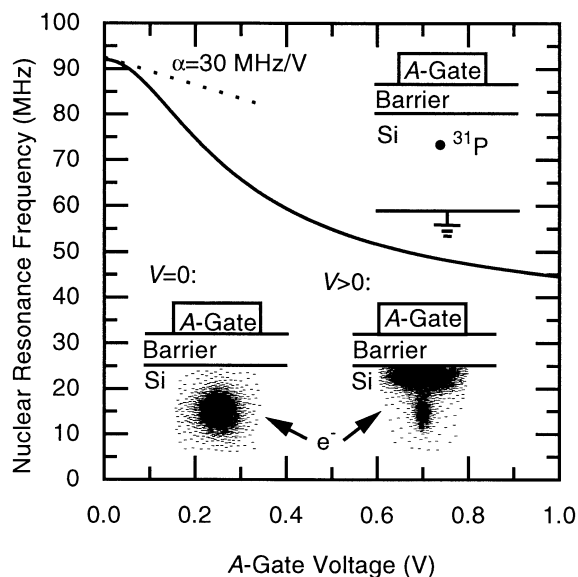


Figure 2 An electric field applied to an A gate pulls the electron wavefunction away from the donor and towards the barrier, reducing the hyperfine interaction and the resonance frequency of the nucleus. The donor nucleus–electron system is a voltage-controlled oscillator with a tuning parameter α of the order of 30 MHz V^{-1} .

required to do elementary operations on the qubits and the separation necessary between donors in the array. The hamiltonian for a nuclear spin–electron system in Si, applicable for an $I = 1/2$ donor nucleus and with $B\|z$ is $H_{\text{en}} = \mu_B B \sigma_z^e - g_n \mu_n B \sigma_z^n + A \sigma^e \cdot \sigma^n$, where σ are the Pauli spin matrices (with eigenvalues ± 1), μ_n is the nuclear magneton, g_n is the nuclear g -factor (1.13 for ^{31}P ; ref. 21), and $A = \frac{8}{3} \pi \mu_B g_n \mu_n |\Psi(0)|^2$ is the contact hyperfine interaction energy, with $|\Psi(0)|^2$, the probability density of the electron wavefunction, evaluated at the nucleus. If the electron is in its ground state, the frequency separation of the nuclear levels is, to second order

$$h\nu_A = 2g_n \mu_n B + 2A + \frac{2A^2}{\mu_B B} \quad (1)$$

In Si: ^{31}P , $2A/h = 58$ MHz, and the second term in equation (1) exceeds the first term for $B < 3.5$ T.

An electric field applied to the electron–donor system shifts the electron wavefunction envelope away from the nucleus and reduces the hyperfine interaction. The size of this shift, following estimates of Kohn²⁵ of shallow donor Stark shifts in Si, is shown in Fig. 2 for a donor 200 Å beneath a gate. A donor nuclear spin–electron system close to an ‘A gate’ functions as a voltage-controlled oscillator: the precession frequency of the nuclear spin is controllable externally, and spins can be selectively brought into resonance with B_{ac} , allowing arbitrary rotations to be performed on each nuclear spin.

Quantum mechanical computation requires, in addition to single spin rotations, the two-qubit ‘controlled rotation’ operation, which rotates the spin of a target qubit through a prescribed angle if, and only if, the control qubit is oriented in a specified direction, and leaves the orientation of the control qubit unchanged^{26,27}. Performing the controlled rotation operation requires nuclear-spin exchange between two donor nucleus–electron spin systems¹³, which will arise from electron-mediated interactions when the donors are sufficiently close to each other. The hamiltonian of two coupled donor nucleus–electron systems, valid at energy scales small compared to the donor–electron binding energy, is $H = H(B) + A_1 \sigma^{1n} \cdot \sigma^{2e} + A_2 \sigma^{2n} \cdot \sigma^{2e} + J \sigma^{1e} \cdot \sigma^{2e}$, where $H(B)$ are the magnetic field interaction terms for the spins. A_1 and A_2 are the hyperfine interaction energies of the respective nucleus–electron systems. $4J$, the exchange energy, depends on the overlap of the electron wavefunctions. For well separated donors²⁸

$$4J(r) \approx 1.6 \frac{e^2}{\epsilon a_B} \left(\frac{r}{a_B} \right)^{\frac{5}{2}} \exp \left(\frac{-2r}{a_B} \right) \quad (2)$$

where r is the distance between donors, ϵ is the dielectric constant of the semiconductor, and a_B is the semiconductor Bohr radius. This function, with values appropriate for Si, is plotted in Fig. 3. Equation (2), originally derived for H atoms, is complicated in Si by its valley degenerate anisotropic band structure²⁹. Exchange coupling terms from each valley interfere, leading to oscillatory behaviour of $J(r)$. In this discussion, the complications introduced by Si band structure will be neglected. In determining $J(r)$ in Fig. 3, the transverse mass for Si ($\approx 0.2m_e$) has been used, and $a_B = 30$ Å. Because J is proportional to the electron wave function overlap, it can be varied by an electrostatic potential imposed by a ‘J-gate’ positioned between the donors¹³. As shall be seen below, significant coupling between nuclei will occur when $4J \approx \mu_B B$, and this condition approximates the necessary separation between donors of 100–200 Å. Whereas actual separations may be considerably larger than this value because the J gate can be biased positively to reduce the barrier between donors, the gate sizes required for the quantum computer are near the limit of current electronics fabrication technology.

For two-electron systems, the exchange interaction lowers the electron singlet ($|\uparrow\downarrow - \downarrow\uparrow\rangle$) energy with respect to the triplets³⁰. (The $|\uparrow\downarrow\rangle$ notation is used here to represent the electron spin state,

and the $|01\rangle$ notation the nuclear state; in the $|\uparrow\downarrow 11\rangle$ state, all spins point in the same direction. For simplicity, normalization constants are omitted.) In a magnetic field, however, $|\uparrow\downarrow\rangle$ will be the electron ground state if $J < \mu_B B/2$ (Fig. 4a). In the $|\uparrow\downarrow\rangle$ state, the energies of the nuclear states can be calculated to second order in A using perturbation theory. When $A_1 = A_2 = A$, the $|10 - 01\rangle$ state is lowered in energy with respect to $|10 + 01\rangle$ by:

$$h\nu_J = 2A^2 \left(\frac{1}{\mu_B B - 2J} - \frac{1}{\mu_B B} \right) \quad (3)$$

The $|\uparrow\downarrow 11\rangle$ state is above the $|10 + 01\rangle$ state and the $|00\rangle$ state below the $|10 - 01\rangle$ state by an energy $h\nu_A$, given in equation (1). For the Si: ^{31}P system at $B = 2$ T and for $4J/h = 30$ GHz, equation (3) yields $\nu_J = 75$ kHz. This nuclear spin exchange frequency approximates the rate at which binary operations can be performed on the computer (ν_J can be increased by increasing J , but at the expense of also increasing the relaxation rate of the coupled nuclear–electron spin excitations). The speed of single spin operations is determined by the size of B_{ac} and is comparable to 75 kHz when $B_{\text{ac}} = 10^{-3}$ T.

Spin measurements

Measurement of nuclear spins in the proposed quantum computer is accomplished in a two-step process: distinct nuclear spin states are adiabatically converted into states with different electron polarization, and the electron spin is determined by its effect on the symmetry of the orbital wavefunction of an exchange-coupled two-electron system. A procedure for accomplishing this conversion is shown in Fig. 4. While computation is done when $J < \mu_B B/2$ and the electrons are fully polarized, measurements are made when $J > \mu_B B/2$, and $|\uparrow\downarrow - \downarrow\uparrow\rangle$ states have the lowest energy (Fig. 4a). As the electron levels cross, the $|\uparrow\downarrow\rangle$ and $|\uparrow\downarrow - \downarrow\uparrow\rangle$ states are coupled by hyperfine interactions with the nuclei. During an adiabatic increase in J , the two lower-energy nuclear spin states at $J = 0$ evolve into $|\uparrow\downarrow - \downarrow\uparrow\rangle$ states when $J > \mu_B B/2$, whereas the two higher-energy nuclear states remain $|\uparrow\downarrow\rangle$. If, at $J = 0$, $A_1 > A_2$, the orientation of nuclear spin 1 alone will determine whether the system evolves into the $|\uparrow\downarrow - \downarrow\uparrow\rangle$ or the $|\uparrow\downarrow\rangle$ state during an adiabatic increase in J .

A method to detect the electron spin state by using electronic

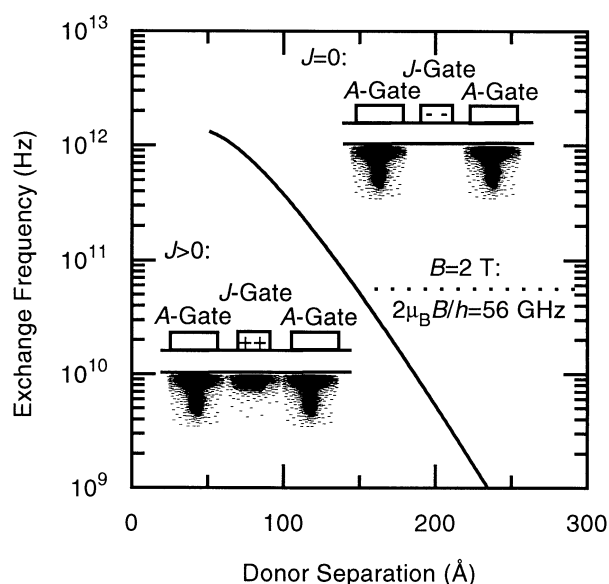


Figure 3 J gates vary the electrostatic potential barrier V between donors to enhance or reduce exchange coupling, proportional to the electron wavefunction overlap. The exchange frequency ($4J/h$) when $V = 0$ is plotted for Si.

means is shown in Fig. 4b. Both electrons can become bound to the same donor (a D^- state) if the A gates above the donors are biased appropriately. In Si:P, the D^- state is always a singlet with a second electron binding energy of 1.7 meV (refs 31, 32). Consequently, a differential voltage applied to the A gates can result in charge motion between the donors that only occurs if the electrons are in a singlet state. This charge motion is measurable using sensitive single-electron capacitance techniques³³. This approach to spin measurement produces a signal that persists until the electron spin relaxes, a time that, as noted above, can be thousands of seconds in Si:P.

The spin measurement process can also be used to prepare nuclear spins in a prescribed state by first determining the state of a spin and flipping it if necessary so that it ends up in the desired spin state. As with the spin computation procedures already discussed, spin measurement and preparation can in principle be performed in parallel throughout the computer.

Initializing the computer

Before any computation, the computer must be initialized by calibrating the A gates and the J gates. Fluctuations from cell to cell in the gate biases necessary to perform logical operations are an inevitable consequence of variations in the positions of the donors and in the sizes of the gates. The parameters of each cell, however, can be determined individually using the measurement capabilities of the computer, because the measurement technique discussed here does not require precise knowledge of the J and A couplings. The A -gate voltage at which the underlying nuclear spin is resonant with an applied B_{ac} can be determined using the technique of adiabatic fast passage³⁴: when $B_{ac} = 0$, the nuclear spin is measured and the A gate is biased at a voltage known to be off resonance. B_{ac} is then switched on, and the A gate bias is swept through a prescribed

voltage interval. B_{ac} is then switched off and the nuclear spin is measured again. The spin will have flipped if, and only if, resonance occurred within the prescribed A -gate voltage range. Testing for spin flips in increasingly small voltage ranges leads to the determination of the resonance voltage. Once adjacent A gates have been calibrated, the J gates can be calibrated in a similar manner by sweeping J -gate biases across resonances of two coupled cells.

This calibration procedure can be performed in parallel on many cells, so calibration is not a fundamental impediment to scaling the computer to large sizes. Calibration voltages can be stored on capacitors located on the Si chip adjacent to the quantum computer. External controlling circuitry would thus need to control only the timing of gate biases, and not their magnitudes.

Spin decoherence introduced by gates

In the quantum computer architecture outlined above, biasing of A gates and J gates enables custom control of the qubits and their mutual interactions. The presence of the gates, however, will lead to decoherence of the spins if the gate biases fluctuate away from their desired values. These effects need to be considered to evaluate the performance of any gate-controlled quantum computer. During the computation, the largest source of decoherence is likely to arise from voltage fluctuations on the A gates. (When $J < \mu_B B/2$, modulation of the state energies by the J gates is much smaller than by the A gates. J exceeds $\mu_B B/2$ only during the measurement process, when decoherence will inevitably occur.) The precession frequencies of two spins in phase at $t = 0$ depends on the potentials on their respective A gates. Differential fluctuations of the potentials produce differences in the precession frequency. At some later time $t = t_\phi$, the spins will be 180° out of phase; t_ϕ can be estimated by determining the transition rate between $|10 + 01\rangle$ (spins in phase) and $|10 - 01\rangle$ (spins 180° out of phase) of a two-spin system. The hamiltonian that couples these states is $H_\phi = \frac{1}{4}h\Delta(\sigma_z^{1n} - \sigma_z^{2n})$, where Δ is the fluctuating differential precession frequency of the spins. Standard treatment of fluctuating hamiltonians³⁴ predicts: $t_\phi^{-1} = \pi^2 S_\Delta(\nu_{st})$, where S_Δ is the spectral density of the frequency fluctuations, and ν_{st} is the frequency difference between the $|10 - 01\rangle$ and $|10 + 01\rangle$ states. At a particular bias voltage, the A gates have a frequency tuning parameter $\alpha = d\Delta/dV$. Thus:

$$t_\phi^{-1} = \pi^2 \alpha^2(V) S_V(\nu_{st}) \quad (4)$$

where S_V is the spectral density of the gate voltage fluctuations.

S_V for good room temperature electronics is of order $10^{-18} \text{ V}^2/\text{Hz}$, comparable to the room temperature Johnson noise of a $50\text{-}\Omega$ resistor. The value of α , estimated from Fig. 2, is $10\text{--}100 \text{ MHz V}^{-1}$, yielding $t_\phi = 10\text{--}1,000 \text{ s}$; α is determined by the size of the donor array cells and cannot readily be reduced (to increase t_ϕ) without reducing the exchange interaction between cells. Because α is a function of the gate bias (Fig. 2), t_ϕ can be increased by minimizing the voltage applied to the A gates.

Although equation (4) is valid for white noise, at low frequencies it is likely that materials-dependent fluctuations ($1/f$ noise) will be the dominant cause of spin dephasing. Consequently, it is difficult to give hard estimates of t_ϕ for the computer. Charge fluctuations within the computer (arising from fluctuating occupancies of traps and surface states, for example) are likely to be particularly important, and minimizing them will place great demands on computer fabrication.

Although materials-dependent fluctuations are difficult to estimate, the low-temperature operations of the computer and the dissipationless nature of quantum computing mean that, in principle, fluctuations can be kept extremely small: using low-temperature electronics to bias the gates (for instance, by using on chip capacitors as discussed above) could produce $t_\phi \approx 10^6 \text{ s}$. Electronically controlled nuclear spin quantum computers thus have the theoretical capability to perform at least 10^5 to perhaps 10^{10}

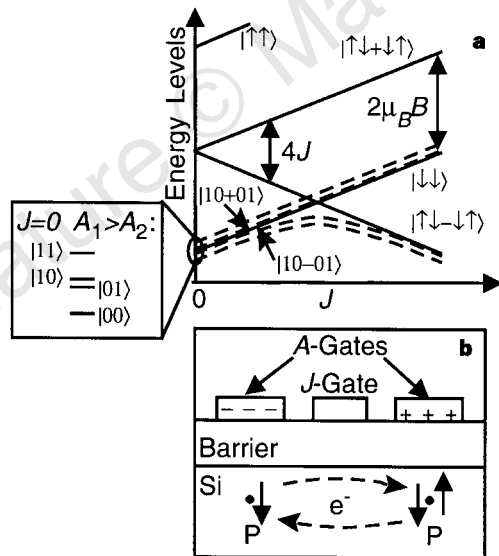


Figure 4 Two qubit quantum logic and spin measurement. **a**, Electron (solid lines) and lowest energy-coupled electron-nuclear (dashed lines) energy levels as a function of J . When $J < \mu_B B/2$, two qubit computations are performed by controlling the $|10 - 01\rangle - |10 + 01\rangle$ level splitting with a J gate. Above $J = \mu_B B/2$, the states of the coupled system evolve into states of differing electron polarization. The state of the nucleus at $J = 0$ with the larger energy splitting (controllable by the A gate bias) determines the final electron spin state after an adiabatic increase in J . **b**, Only $|\uparrow\uparrow\rangle - |\downarrow\downarrow\rangle$ electrons can make transitions into states in which electrons are bound to the same donor (D^- states). Electron current during these transitions is measurable using capacitive techniques, enabling the underlying spin states of the electrons and nuclei to be determined.

logical operations during t_ϕ , and can probably meet Preskill's criterion⁸ for an error probability of 10^{-6} per qubit operation.

Constructing the computer

Building the computer presented here will obviously be an extraordinary challenge: the materials must be almost completely free of spin ($I \neq 0$ isotopes) and charge impurities to prevent dephasing fluctuations from arising within the computer. Donors must be introduced into the material in an ordered array hundreds of Å beneath the surface. Finally, gates with lateral dimensions and separations ~ 100 Å must be patterned on the surface, registered to the donors beneath them. Although it is possible that the computer can use SiO_2 as the barrier material (the standard MOS technology used in most current conventional electronics), the need to reduce disorder and fluctuations to a minimum means that heteroepitaxial materials, such as Si/SiGe , may ultimately be preferable to Si/SiO_2 .

The most obvious obstacle to building to the quantum computer presented above is the incorporation of the donor array into the Si layer beneath the barrier layer. Currently, semiconductor structures are deposited layer by layer. The δ -doping technique produces donors lying on a plane in the material, with the donors randomly distributed within the plane. The quantum computer envisaged here requires that the donors be placed into an ordered one- or two-dimensional array; furthermore, precisely one donor must be placed into each array cell, making it extremely difficult to create the array by using lithography and ion implantation or by focused deposition. Methods currently under development to place single atoms on surfaces using ultra-high-vacuum scanning tunnelling microscopy³⁵ or atom optics techniques³⁶ are likely candidates to be used to position the donor array. A challenge will be to grow high-quality Si layers on the surface subsequent to placement of the donors.

Fabricating large arrays of donors may prove to be difficult, but two-spin devices, which can be used to test the logical operations and measurement techniques presented here, can be made using random doping techniques. Although only a small fraction of such devices will work properly, adjacent conventional Si electronic multiplexing circuitry can be used to examine many devices separately. The relative ease of fabricating such 'hybrid' (quantum-conventional) circuits is a particularly attractive feature of Si-based quantum computation.

In a Si-based nuclear spin quantum computer, the highly coherent quantum states necessary for quantum computation are incorporated into a material in which the ability to implement complex computer architectures is well established. The substantial challenges facing the realization of the computer, particularly in fabricating 100-Å-scale gated devices, are similar to those facing the next generation of conventional electronics; consequently, new manufacturing technologies being developed for conventional electronics will bear directly on efforts to develop a quantum computer in Si. Quantum computers sufficiently complex that they

can achieve their theoretical potential may thus one day be built using the same technology that is used to produce conventional computers. □

Received 10 November 1997; accepted 24 February 1998.

1. Steane, A. Quantum computing. *Rep. Prog. Phys.* **61**, 117–173 (1998).
2. Bennett, C. H. Quantum information and computation. *Physics Today* 24–30 (Oct. 1995).
3. Shor, P. W. in *Proc. 35th Annu. Symp. Foundations of Computer Science* (ed. Goldwasser, S.) 124–134 (IEEE Computer Society, Los Alamitos, CA, 1994).
4. Ekert, A. & Jozsa, R. Quantum computation and Shor's factoring algorithm. *Rev. Mod. Phys.* **68**, 733–753 (1996).
5. Grover, L. K. Quantum mechanics helps in searching for a needle in a haystack. *Phys. Rev. Lett.* **79**, 325–328 (1997).
6. Calderbank, A. R. & Shor, P. W. Good quantum error correcting codes exist. *Phys. Rev. A* **54**, 1098–1105 (1996).
7. Steane, A. M. Error correcting codes in quantum theory. *Phys. Rev. Lett.* **77**, 793–797 (1996).
8. Preskill, J. Reliable quantum computers. *Proc. R. Soc. Lond. A* **454**, 385–410 (1998).
9. Lloyd, S. A potentially realizable quantum computer. *Science* **261**, 1569–1571 (1993).
10. DiVincenzo, D. P. Quantum computation. *Science* **270**, 255–261 (1995).
11. Gershenfeld, N. A. & Chuang, I. L. Bulk spin-resonance quantum computation. *Science* **275**, 350–356 (1997).
12. Cory, D. G., Fahmy, A. F. & Havel, T. F. Ensemble quantum computing by NMR spectroscopy. *Proc. Natl Acad. Sci. USA* **94**, 1634–1639 (1997).
13. Loss, D. & DiVincenzo, D. P. Quantum computation with quantum dots. *Phys. Rev. A* **57**, 120–126 (1998).
14. Privman, V., Vagner, I. D. & Kventzel, G. Quantum computation in quantum Hall systems. *Phys. Lett. A* **239**, 141–146 (1998).
15. Slichter, C. P. *Principles of Magnetic Resonance* 3rd edn, Ch 4 (Springer, Berlin, 1990).
16. Dobers, M., Klitzing, K. v., Schneider, J., Weimann, G. & Ploog, K. Electrical detection of nuclear magnetic resonance in $\text{GaAs-AlGa}_{1-x}\text{As}$ heterostructures. *Phys. Rev. Lett.* **61**, 1650–1653 (1988).
17. Stich, B., Greulich-Weber, S. & Spaeth, J.-M. Electrical detection of electron nuclear double resonance in silicon. *Appl. Phys. Lett.* **68**, 1102–1104 (1996).
18. Kane, B. E., Pfeiffer, L. N. & West, K. W. Evidence for an electric-field-induced phase transition in a spin-polarized two-dimensional electron gas. *Phys. Rev. B* **46**, 7264–7267 (1992).
19. Wald, K. W., Kouwenhoven, L. P., McEuen, P. L., van der Vaart, N. C. & Foxon, C. T. Local dynamic nuclear polarization using quantum point contacts. *Phys. Rev. Lett.* **73**, 1011–1014 (1994).
20. Dixon, D. C., Wald, K. R., McEuen, P. L. & Melloch, M. R. Dynamic polarization at the edge of a two-dimensional electron gas. *Phys. Rev. B* **56**, 4743–4750 (1997).
21. *CRC Handbook of Chemistry and Physics* 77th edn 11–38 (CRC Press, Boca Raton, Florida, 1996).
22. Feher, G. Electron spin resonance on donors in silicon. I. Electronic structure of donors by the electron nuclear double resonance technique. *Phys. Rev.* **114**, 1219–1244 (1959).
23. Wilson, D. K. & Feher, G. Electron spin resonance experiments on donors in silicon. III. Investigation of excited states by the application of uniaxial stress and their importance in relaxation processes. *Phys. Rev.* **124**, 1068–1083 (1961).
24. Waugh, J. S. & Slichter, C. P. Mechanism of nuclear spin-lattice relaxation in insulators at very low temperatures. *Phys. Rev. B* **37**, 4337–4339 (1988).
25. Kohn, W. *Solid State Physics* Vol. 5 (eds Seitz, F. & Turnbull, D.) 257–320 (Academic, New York, 1957).
26. DiVincenzo, D. P. Two-bit gates are universal for quantum computation. *Phys. Rev. A* **51**, 1015–1021 (1995).
27. Lloyd, S. Almost any quantum logic gate is universal. *Phys. Rev. Lett.* **75**, 346–349 (1995).
28. Herring, C. & Flicker, M. Asymptotic exchange coupling of two hydrogen atoms. *Phys. Rev.* **134**, A362–A366 (1964).
29. Andres, K., Bhatt, R. N., Goalwin, P., Rice, T. M. & Walstedt, R. E. Low-temperature magnetic susceptibility of Si:P in the nonmetallic region. *Phys. Rev. B* **24**, 244–260 (1981).
30. Ashcroft, N. W. & Mermin, N. D. in *Solid State Physics* Ch. 32 (Saunders College, Philadelphia, 1976).
31. Larsen, D. M. Stress dependence of the binding energy of D^+ centers in Si. *Phys. Rev. B* **23**, 5521–5526 (1981).
32. Larsen, D. M. & McCann, S. Y. Variational studies of two- and three-dimensional D^+ centers in magnetic fields. *Phys. Rev. B* **46**, 3966–3970 (1992).
33. Ashoori, R. C. Electrons in artificial atoms. *Nature* **379**, 413–419 (1996).
34. Abragam, A. *Principles of Nuclear Magnetism* (Oxford Univ. Press, London, 1961).
35. Lyding, J. W. UHV STM nanofabrication: progress, technology spin-offs, and challenges. *Proc. IEEE* **85**, 589–600 (1997).
36. Adams, C. S., Sigel, J. & Mlynek, J. Atom optics. *Phys. Rep.* **240**, 143–210 (1994).
37. Warren, W. S. The usefulness of NMR quantum computing. *Science* **277**, 1688–1690 (1997).

Acknowledgements. This work has been supported by the Australian Research Council. I thank R. G. Clark for encouragement and E. Hellman for suggesting that the work in ref. 18 could be relevant to quantum computation.

Correspondence should be addressed to the author (e-mail: kane@newt.phys.unsw.edu.au).

Birth kicks as the origin of pulsar rotation

H. Spruit* & E. S. Phinney†‡

* Max-Planck-Institut für Astrophysik, Postfach 1523, D-85740 Garching, Germany

† European Southern Observatory, Karl-Schwarzschild-Strasse 2, Garching bei München, Germany

‡ California Institute of Technology, Theoretical Astrophysics 130-33, Pasadena, California 91125, USA

Radio pulsars are thought to be born with spin periods of 0.02–0.5 s and space velocities of 100–1,000 km s^{−1}, and they are inferred to have initial dipole magnetic fields of 10¹¹–10¹³ G (refs 1–5). The average space velocity of their progenitor stars is less than 15 km s^{−1}, which means that pulsars must receive a substantial ‘kick’ at birth. Here we propose that the birth characteristics of pulsars have a simple physical connection with each other. Magnetic fields maintained by differential rotation between the core and envelope of the progenitor would keep the whole star in a state of approximately uniform rotation until 10 years before the explosion. Such a slowly rotating core has 1,000 times less angular momentum than required to explain the rotation of pulsars. The specific physical process that ‘kicks’ the neutron star at birth has not been identified, but unless its force is exerted exactly head-on⁶ it will also cause the neutron star to rotate. We identify this process as the origin of the spin of pulsars. Such kicks may cause a correlation between the velocity and spin vectors of pulsars. We predict that many neutron stars are born with periods longer than 2 s, and never become radio pulsars.

A common interpretation of the magnetic fields and spins of pulsars appeals to the magnetic fields and rotation of the main-sequence stars from which neutron stars are derived⁷. Compressing the Sun to the size of a neutron star (hypothetically, as the Sun is not a supernova progenitor), conserving (surface dipole) magnetic flux and angular momentum, would produce a neutron star of surface dipole field $B_s \approx 10^{11}$ G and spin period $P \approx 0.03$ s, reminiscent of observed values. For this explanation to work, the small core of the supernova progenitor must be able to spin freely in its large, slowly rotating envelope.

The magnetic fields inside stars are unobservable, but rotation is, in the case of the Sun. The rotation below the convective envelope is essentially uniform^{8,9}, with measured degrees of differential rotation well below the 30% level seen at the solar surface. The spin-down torque due to the solar wind has apparently been transmitted very effectively to the core. Such a low degree of differential rotation is incompatible¹⁰ with the hydrodynamic angular-momentum transport mechanisms commonly used in stellar evolution calculations¹¹.

Simple estimates¹² and detailed calculations¹³ show that the Maxwell stresses should be sufficient to maintain a state of approximately uniform rotation in a star as long as it has at least a weak (~ 1 G) initial magnetic field. If this holds also for pre-supernovae, their cores do not spin freely. We now explore the consequences of assuming effective magnetic coupling between cores and envelopes of stars. With current understanding of magnetic fields and the evidence provided by the Sun, this assumption is not less plausible than the currently more common view that evolved stars have rapidly rotating cores.

The magnetic field strength transmitting torques is conveniently measured by the quantity $\bar{B} \equiv (B_r B_\theta)^{1/2}$, the geometric mean of the radial and toroidal field components, approximately the minimum total field required for a fixed stress. An initially weak field in a differentially rotating star is quickly amplified, by field line stretch-

ing, instabilities^{14,15}, and perhaps also dynamo action^{16,17}, until the Maxwell stresses are sufficient to enforce a nearly uniform rotation throughout the star. During most of its life, the star’s rotation period is very short compared with the timescale τ_c on which the moment of inertia $I_c = k M_c R_c^2$ of the core evolves, and the fields can maintain nearly uniform rotation at angular frequency Ω throughout the star. At the edge of the core the required field is:

$$\bar{B} \approx \left(\frac{I_c \Omega}{R_c^3 \tau_c} \right)^{1/2} \quad (1)$$

where M_c , R_c are respectively the core’s mass and radius, and $k \approx 0.2$. In the short-lived late stages of nuclear burning, however, the situation reverses. Trying to remain locked to a slowly rotating supergiant envelope, the core’s rotation period becomes longer than the increasingly short timescale on which it is shrinking. Equation (1) implies that the magnetic energy in the fields required to keep the core co-rotating with the envelope is at least of the order of the total rotational energy in the core, divided by $\Omega \tau_c$. As differential rotation supplies the free energy to build up the fields, when $\Omega \tau_c < 1$, there is no longer enough free energy to grow magnetic fields large enough for Maxwell stresses to keep the core in co-rotation. As long as $\Omega \tau_c > 1$, the core co-rotates, but once the core’s contraction time drops below $1/\Omega$, it decouples and the further contraction takes place under approximate conservation of angular momentum. From more detailed calculations, we find that, for an isolated star of 15 solar masses which began life on the main sequence with the 200 km s^{−1} rotation speed typical of early B stars¹⁸, this decoupling happens no later than carbon depletion, when the core evolution timescale is ~ 4 yr, the core size 10⁹ cm, and the core field strength $\bar{B} \approx 10^4$ G. The field strength is increased somewhat by the differential rotation but largely by compression. Integrating the further evolution, we find for the newborn neutron star a rotation period $P_0 \approx 100$ s and a field strength $\bar{B} \approx 10^{11}$ G.

Young pulsars have dipole magnetic fields in the range of these minimum fields, but their observed rotation periods are 3–4 orders of magnitude shorter than our prediction for the cores of isolated stars. As we now argue, instead of being a remnant of the initial stellar rotation, the birth spins of pulsars are probably imparted by the same events that give them their velocities.

Consider a young neutron star of mass M , and radius $R = f_\Omega R_{NS}$, where R_{NS} is the final (~ 10 km) cooled radius and $f_\Omega \approx 2$ –3 during the bulk of the neutrino emission. The final moment of inertia $I_{NS} = k M R_{NS}^2$, where $k \approx 0.36$. If momentum impulses $\Delta \mathbf{P}_i$ are applied at radii $\mathbf{R}_i$ to the young neutron star, the final velocity is $\mathbf{v} = M^{-1} \Sigma_i \Delta \mathbf{P}_i$ and the final angular velocity is given by $\boldsymbol{\Omega}_{NS} = \Delta \mathbf{I} / I_{NS} = I_{NS}^{-1} \Sigma_i \mathbf{R}_i \times \Delta \mathbf{P}_i$. Unless the impulses which give the neutron star its space velocity always have an exactly zero angle α_i between the direction of the impulse and the radius vector from the centre of mass, the kicks will impart spin, for the same reason that kicked footballs and flipped coins spin. If all neutron stars received their velocities from a single localized momentum impulse, they would always rotate about an axis perpendicular to their direction of motion⁶. If all impulses also occurred at a fixed impact parameter $d = R \sin \alpha$, then $\Omega \propto v$: faster moving pulsars would spin faster. The final spin in this oversimplified case is related to the kick speed by $\Omega_{NS} = (f_\Omega/k)(v/R_{NS}) \sin \alpha$, or numerically as a period:

$$P_{NS} = 0.07 \left(\frac{200 \text{ km s}^{-1}}{v} \right) \left(\frac{0.5}{\sin \alpha} \right) \left(\frac{3}{f_\Omega} \right) \text{ s} \quad (2)$$

But if the momentum impulses are due to convection^{6,19–21} (leading to anisotropic neutrino transport, or perhaps also anisotropic fall-back^{22,23}), then more than one impulse will probably be applied, with directions distributed about the local radial directions. It is then possible to impart spin with no space velocity, and space velocity with no spin. Indeed, if many small independent momen-

tum impulses are applied at random positions on the neutron-star surface, in directions with a distribution azimuthally symmetric about the local radial direction, the resulting kick velocity and angular velocity are uncorrelated in both magnitude and direction, with both quantities having a maxwellian distribution. The tendency for the spin axis to be perpendicular to the velocity is destroyed especially rapidly. Four or more kicks produce a nearly uniform distribution of $\mathbf{v} \cdot \boldsymbol{\Omega}$.

To quantify this, we introduce an (arbitrary) z -axis, and let impulse $\Delta \mathbf{P}_i$ be applied at the surface radius R at a point at polar angle θ_i , at an angle α_i to the local radial direction; the angle between the $\mathbf{z}$ - $\mathbf{P}_i$ plane and the $\mathbf{R}_i$ - $\mathbf{P}_i$ plane being ϕ_i . Then $\Omega_z = -I_{\text{NS}}^{-1} \sum_i \Delta P_i \sin \theta_i \sin \alpha_i \sin \phi_i$, and $v_z = M^{-1} \sum_i \Delta P_i \cos \theta_i$, where $\cos \theta_i = \cos \theta_i \cos \alpha_i - \sin \theta_i \sin \alpha_i \cos \phi_i$. If θ_i is uniform over the sphere, ϕ_i uniform on $[0, 2\pi]$, and the α_i are drawn from a position-independent distribution, then $\langle v_z \rangle = \langle \Omega_z \rangle = \langle v_z \Omega_z \rangle = \langle v_z \Omega_z \rangle = 0$, and $\langle v_z^2 \rangle = \frac{1}{3} M^{-2} \sum_i \Delta P_i^2$, while $\langle \Omega_z^2 \rangle = \frac{1}{3} R^2 I_{\text{NS}}^{-2} \langle \sin^2 \alpha \rangle \sum_i \Delta P_i^2$. Thus

$$\langle \Omega_z^2 \rangle^{1/2} = \langle \sin^2 \alpha \rangle^{1/2} \frac{f_{\Omega} \langle v_z^2 \rangle^{1/2}}{k R_{\text{NS}}} \quad (3)$$

and similarly for the uncorrelated x and y components. This expression is the same as for single impulses, but with individual quantities replaced by r.m.s. ones. The distribution of pulsar birth velocities inferred from observed proper motions can be described⁵ by a maxwellian with one-dimensional dispersion $\langle v_z^2 \rangle^{1/2} = 190 \text{ km s}^{-1}$, though other distributions marginally fit. Inserting this number into equation (3) gives the predicted r.m.s. rotation rate at birth. The corresponding period is $\bar{P} = 0.08(0.5/\langle \sin^2 \alpha \rangle^{1/2})(3/f_{\Omega})$.

In the above discussion, we took the momentum impulses to be instantaneous. This is a good approximation if the duration of the kicks is short compared with the resulting rotation period. The timescale of the processes^{19,20,21,24} relevant for the kick ranges from 0.01 s to several seconds, and thus may have to be taken into account. If the asymmetries are fixed in a frame rotating with the young neutron star, and thrusts are active at several locations on the surface, rotational averaging will reduce the contribution to the kick velocity perpendicular to the rotation axis, but not the contribution to the angular velocity. The contribution to the star's linear

momentum from components of the thrusts along the rotation axis will build up linearly with time, while the components in the two traverse directions will rotationally average away. If P_0 is the final rotation period and τ the duration of the thrust, this happens if $P_0 < (1/9)\tau$ (for $f_{\Omega} = 3$). In this limit, the birth velocity will therefore be preferentially aligned with the rotation axis, in contrast to the case of a single thrust, where the velocity and spin vectors must be precisely orthogonal. If the thrust duration is comparable to the induced rotation periods, shorter-period systems will have significant rotational averaging. If there are several thrusts active at random locations, short-period pulsars have mean birth velocities lower by $\sim 1/\sqrt{3}$ compared to longer-period systems. If only a single long-duration thrust acts, the transverse velocity can be reduced by even larger factors compared with the instantaneous kick case.

Data from some 60 pulsars with known transverse velocities are shown in Fig. 1, and compared with a computation for parameters which approximately reproduce the observations. In this calculation, four thrusts are applied for a duration $\tau = 0.32 \text{ s}$. As pulsars spin down significantly by emission of electromagnetic radiation, the initial spin periods can not be compared directly with observations. The right panel of Fig. 1 shows the result of applying spin-down and selection effects according to the recipes of Bhattacharya *et al.*². The periods now pile up towards the pulsar 'death line' (the period, near 1 s, beyond which pulsars stop producing radio emission), and the correlation between spin and velocity is washed out. The model has a modest correlation between the directions of spin and velocity (preference for alignment). We note that the data seem to contain fewer high-velocity short-period pulsars than the simulation. In part, this is due to the larger density of points used in the simulated distributions (containing ~ 300 pulsars). But it may also be a possible indication for long-duration kicks ($\sim 1 \text{ s}$) acting at a single 'hot spit' on the proto-neutron star. This case gives a stronger rotational averaging effect of the space velocity, at short periods, than the example shown.

Though our initial minimum $\bar{B}$ is comparable to the observed dipole field strengths of young pulsars, there are further processes that can modify the field strength. The vigorous convection in the newly formed neutron star will tangle the interior field lines. As the field strength in equipartition with the convective motions is $\sim 10^{15} \text{ G}$ (ref. 25), this phase could substantially modify the fields.

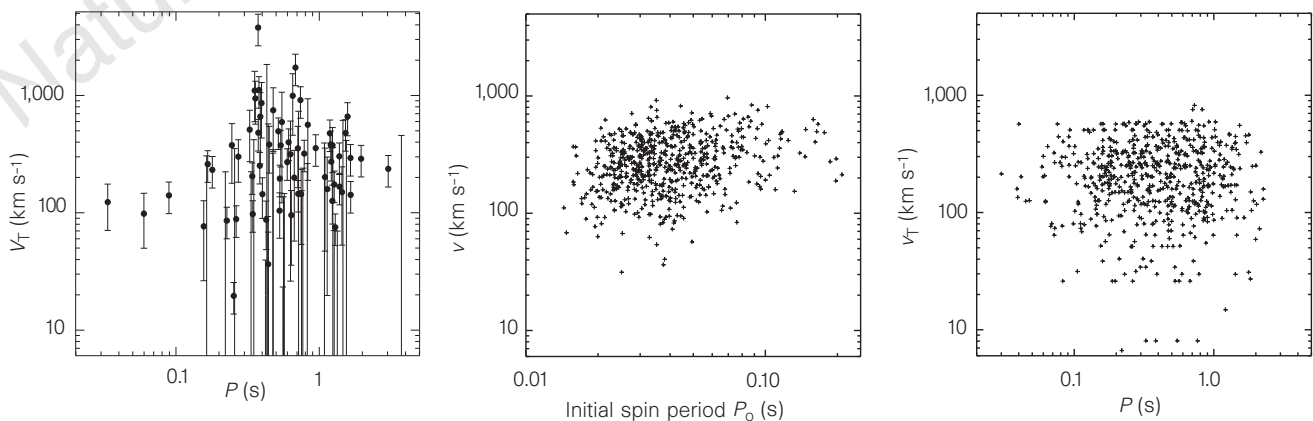


Figure 1 Rotation periods and velocities of pulsars compared with model. Left panel, observed periods, and transverse velocities of radio pulsars. Middle panel, Birth periods and space velocities of 'kicked' pulsars. Four thrusts of $T_i = 1.8 \times 10^{41} \text{ dyn}$ are applied at positions randomly located on the surface of a 30-km ($f_{\Omega} = 3$) proto-neutron star at random angles with 45° of the inward normal ($\langle \sin^2 \alpha \rangle = 0.26$), for a duration $\tau = 0.32 \text{ s}$. For this thrust duration the velocity components perpendicular to the rotation axis are significantly reduced, for the shorter spin periods. For an impulse of thrust T and duration τ applied at the surface $R = f_{\Omega} R_{\text{NS}}$ of a star initially non-rotating and at rest, the final angular velocity is the

same as for an instantaneous impulse $T\tau$: $\Omega_{\text{NS}} = (T\tau)R \sin \alpha / I_{\text{NS}}$. But the velocity imparted to the neutron star is lower: $v = M^{-1}(\pi f_{\Omega} / \pi) T / (R_{\text{NS}} \sin \alpha)^{1/2} L(s)$, where $s = f_{\Omega}^{-1}(\tau \Omega_{\text{NS}} / \pi)^{1/2}$, $L(s) = (C(s) + S(s))^{1/2} < s$, and $C(s)$ and $S(s)$ are the Fresnel cosine and sine integrals, respectively. For this reason, the highest velocities do not occur at the shortest periods in the model shown. Right panel, same model, but taking into account spin-down, and showing only the observable (transverse) velocity component. Pulsars are assumed to be born uniformly distributed in space; spin-down and observational selection effects according to Bhattacharya *et al.*².

We have been deliberately vague about the source, number and duration of the natal thrusts, as current models vary significantly. The magnitude of the initial spins (equations (2), (3)) we predict are insensitive to the nature of the kick mechanism. But the distributions, and particularly the spin-velocity correlation are sensitive to it. Recent simulations^{24,26} have shown that the interior of a young neutron star is convective for at least 1 s, during which time some 10^4 convection cells come and go (making the impulse treatment above the relevant one). However, the neutrino flux varies by only $\sim 4\%$ from isotropy on the scale of the cells, so the estimated recoil velocity is only $\sim 0.04(10^4)^{-1/2}GM/(R_{\text{NSC}}) \sim 30 \text{ km s}^{-1}$. Attaining the observed velocities may require more global hydrodynamic, or magnetically imposed, asymmetries. A promising possibility is the effect of parity violation on neutrino scattering in a magnetic field^{27,28}. This effect would produce long-duration (several seconds) thrusts. Assuming that this is the correct mechanism, our calculations then predict that rotation axis and transverse velocity should tend to be along the same direction on the sky. The magnitudes of velocity and (initial) spin will correlate only weakly, because the thrust component that produces the spin is also the one whose contribution to the space motion gets 'washed out' by the rotation. Independent of the precise kick mechanism, however, our model predicts a wide distribution of initial spin periods. Many of these spin periods may be long enough ($>2 \text{ s}$) to ensure that radio emission is not produced. We consider that the recently discovered^{29,30} young, slowly rotating isolated neutron stars (detected by the Rosat and ASCA satellites) are such cases. $\square$

Received 7 October 1997; accepted 12 March 1998

- Narayan, R. The birthrate and initial spin period of single radio pulsars. *Astrophys. J.* **319**, 162–179 (1987).
- Bhattacharya, D., Wijers, R. A. M. J., Hartman, J. W. & Verbunt, F. On the decay of the magnetic fields of single radio pulsars. *Astron. Astrophys.* **254**, 198–212 (1992).
- Lorimer, D. R., Bailes, M., Dewey, R. J. & Harrison, P. A. Pulsar statistics—The birth rates and initial spin periods of radio pulsars. *Mon. Not. R. Astron. Soc.* **263**, 403–415 (1993).
- Lyne, A. G. & Lorimer, D. R. High birth velocities of radio pulsars. *Nature* **369**, 127–129 (1994).
- Hansen, B. M. S. & Phinney, E. S. The pulsar kick velocity distribution. *Mon. Not. R. Astron. Soc.* **291**, 569–577 (1997).
- Burrows, A., Hayes, J. & Fryxell, B. A. On the nature of core-collapse supernova explosions. *Astrophys. J.* **450**, 830–850 (1995).
- Langer, N., Fliegner, J., Heger, A. & Woosley, S. E. Nucleosynthesis in rotating massive stars. *Nucl. Phys. A* **621**, 475c–466c (1997).
- Thompson, M. J. *et al.* Solar internal rotation and dynamics from GONG frequency splittings. *Science* **272**, 1300–1305 (1996).
- Corbard, T. *et al.* Solar internal rotation from LOWL data. A 2D regularized least-squares inversion using B-splines. *Astron. Astrophys.* **324**, 298–310 (1997).
- Spruit, H. C., Knoblock, E. & Roxburgh, I. W. Internal rotation of the sun. *Nature* **304**, 520–522 (1983).
- Endal, A. S. & Sofia, S. The evolution of rotating stars. II—Calculations with time-dependent redistribution of angular momentum for 7- and 10-solar-mass stars. *Astrophys. J.* **220**, 279–290 (1978).
- Mestel, L. Rotation and stellar evolution. *Mon. Not. R. Astron. Soc.* **113**, 735 (1953).
- Charbonneau, P. & MacGregor, K. B. Angular momentum transport in magnetized stellar radiative zones. I—Numerical solutions to the core spin-up model problem. *Astrophys. J.* **387**, 639–661 (1992).
- Acheson, D. J. On the instability of toroidal magnetic fields and differential rotation in stars. *Phil. Trans. R. Soc. Lond.* **289**, 459–500 (1978).
- Balbus, S. A. General local stability criteria for stratified, weakly magnetized rotating systems. *Astrophys. J.* **453**, 380–383 (1995).
- Tout, C. A. & Pringle, J. E. accretion disc velocity—A simple model for a magnetic dynamo. *Mon. Not. R. Astron. Soc.* **256**, 269–276 (1992).
- Regös, E. & Tout, C. A. The effect of magnetic fields in common-envelope evolution on the formation of cataclysmic variables. *Mon. Not. R. Astron. Soc.* **273**, 146–156 (1995).
- Slettebak, A. in *Stellar Rotation* (ed. Slettebak, A.) 3–8 (Reidel, Dordrecht, 1970).
- Herant, M., Benz, W. & Colgate, S. Postcollapse hydrodynamics of SN 1987A—Two-dimensional simulations of the early evolution. *Astrophys. J.* **395**, 642–653 (1992).
- Burrows, A. & Hayes, J. Pulsar recoil and gravitational radiation due to asymmetrical stellar collapse and explosion. *Phys. Rev. Lett.* **76**, 352–355 (1996).
- Janka, H.-Th. & Müller, E. Neutrino heating, convection, and the mechanism of Type-II supernova explosions. *Astron. Astrophys.* **306**, 167–198 (1996).
- Janka, H.-Th. & Müller, E. Neutron star recoils from anisotropic supernovae. *Astron. Astrophys.* **290**, 496–502 (1994).
- Janka, H.-Th. & Müller, E. Neutron star kicks and multi-dimensional supernova models. *Ann. NY Acad. Sci.* **759**, 269–274 (1995).
- Keil, W., Janka, H.-Th. & Müller, E. Ledoux convection in protoneutron stars—clue to supernova nucleosynthesis? *Astrophys. J.* **473**, L111–L114 (1996).
- Thompson, C. & Duncan, R. C. Neutron star dynamos and the origins of pulsar magnetism. *Astrophys. J.* **408**, 194–217 (1993).
- Burrows, A. Convection and the mechanism of type II supernovae. *Astrophys. J.* **318**, L57–L61 (1987).
- Horowitz, C. J. & Li, G. Cumulative parity violation in supernovae. Preprint hep-ph/9701214 (1997).
- Lai, D. & Qian, Y.-Z. Neutrino transport in strongly magnetized proto-neutron stars and the origin of pulsar kicks: I. The effect of parity violation in weak interactions. *Astrophys. J.* (submitted): preprint astro-ph/9802344 (1998).

- Haberl, F., Motch, C., Buckley, D. A. H., Zickgraf, F.-J. & Pietsch, W. RXJ0720.4–3125: strong evidence for an isolated pulsating neutron star. *Astron. Astrophys.* **326**, 662–668 (1997).
- Mergehetti, S., Stella, L. & Israel, G. L. Recent results on anomalous X-ray pulsars. Preprint astro-ph/9712254 (1997).

Acknowledgements. E.S.P. thanks G. Soberman and S. Woosley for the sequences of stellar models used in the integrations. E.S.P. is supported in part by the US NSF and NASA, and thanks ESO for hospitality and support.

Correspondence and requests for materials should be addressed to H.C.S. (e-mail: henk@mpa-garching.mpg.de).

Astrometric signatures of giant-planet formation

Alan P. Boss

Department of Terrestrial Magnetism, Carnegie Institution of Washington, 5241 Broad Branch Road, NW, Washington DC 20015-1305, USA

The discovery^{1–6} of giant planets orbiting nearby solar-type stars raises anew the question of how they formed. Two very different mechanisms have been proposed. Gravitational instability^{7,8} is the process by which planets form directly from the gas in the accretion disk around the young star. The other alternative is core accretion^{9,10}, where rocky cores of about 10 Earth masses form, followed by the hydrodynamical accretion of gas. Here I show that these processes have very different astrometric signatures, and that it is observationally possible to distinguish between them. Planets that form through gravitational instabilities do so rapidly, so that within just a few hundred years of the onset of the instability the nascent planet is making the young stellar object wobble in its orbit. This can be seen in the youngest stellar objects, with ages as little as 0.1 Myr. If planets form by core accretion, an observable wobble will not be visible for 10–20 Myr. Observations of a suitable ensemble of optically visible young stellar objects (such as those in the Taurus molecular cloud) over a period of several decades should be able to determine which of the two processes is responsible for giant-planet formation.

Whereas there is widespread agreement that the terrestrial planets formed over a time period of ~ 100 Myr by the collisional accumulation of successively larger planetesimals and planetary embryos¹¹, the formation of the giant planets (Jupiter and Saturn) is more controversial. In the core-accretion model, giant-planet formation cannot occur until planetary embryos of ~ 10 Earth masses ($10 M_{\oplus}$) form, which apparently required at least 1 Myr for our Solar System¹². The subsequent accretion of up to $\sim 300 M_{\oplus}$ of nebular gas to form the massive envelopes of the giant planets takes almost 10 Myr in the most recent models¹². Both of these times are relative to the epoch when ~ 100 -km planetesimals have formed, that is, after the protostar has formed and the nebula has become quiescent enough to allow dust grain growth to planetesimal size. Core accretion could start to occur around any young stellar object (YSO) which has a quiescent, moderately massive disk, conceivably around even the youngest YSOs. The age of the star at the epoch when Jupiter-mass planets form will then be the sum of the stellar age when core accretion started (0.1–10 Myr) plus the time needed for the core-accretion mechanism to occur, 10 Myr. In the core-accretion model, then, astrometric wobbles caused by Jupiter-mass protoplanets should not appear before YSO ages of about 10–20 Myr. Before this age, only Earth- to Uranus-mass protoplanets could exist in the giant-planet formation region.

The disk-instability mechanism, on the other hand, can lead to the formation of multiple-Jupiter-mass protoplanets within a few hundred years after the disk becomes gravitationally unstable⁸. Gravitational instability requires a relatively massive, cool disk, a situation that may occur early in the evolution of protostellar disks¹³. Gravitational torques associated with, for example, spiral

arms will help drive the inward transport of the disk's mass onto the protostar¹⁴. However, giant gaseous protoplanets formed during this phase might be swallowed by the growing protosun. To avoid this fate, giant gaseous protoplanets must form towards the end of the main protostellar accretion phase, when most of the disk mass has already been added to the protostar, yet the disk is still massive enough to be unstable. Disk instability probably requires a YSO as young as ~ 0.1 –1 Myr. Because the gravitational instability mechanism is essentially instantaneous compared to these ages, the stellar ages at which giant planets would form through this mechanism also range from ~ 0.1 to 1 Myr. Studies of the orbital migration of giant planets through gravitational interactions with the disk suggest that inward orbital migration occurs on timescales of several to ~ 10 Myr, depending on the mass of the planet, with more massive planets migrating much more slowly¹⁵. Massive planets orbiting young (~ 0.1 –1 Myr) YSOs therefore would not have had sufficient time to suffer significant orbital decay.

If young (~ 0.1 –1 Myr) YSOs wobble significantly, then the core accretion mechanism of giant-planet formation can be ruled out, at least as it is presently understood. Giant-planet formation would have to occur by gravitational instability in this case. On the other hand, if the youngest YSOs do not wobble, but the oldest (~ 10 Myr) YSOs do wobble, then core accretion would be the favoured mechanism. If even the oldest, now-diskless YSOs do not wobble, then either giant-planet formation is not taking place at all, which would be highly unlikely given that we now know that extrasolar giant planets do exist, or giant-planet formation does not take place in circumstellar disks, an almost inconceivable outcome. A sample size of the order of 100 optically visible YSOs of varying ages would probably be necessary to provide an unambiguous test of these formation mechanisms. Nearby star-forming regions such as Taurus-Auriga and ρ Ophiuchi contain sufficient numbers of optically visible young stars.

The astrometric method of planet detection¹⁶ infers the presence of unseen companions by measuring the position of the primary star as it orbits around the centre of mass of the entire system. Astrometric observations have suggested the presence of one or more Jupiter-mass planets in orbit around the nearby star Lalande 21185 (ref. 1). Most extrasolar planets have been discovered by another indirect method^{2–6}, the spectroscopic method, where the star's orbit about the centre of mass is deduced from the periodic Doppler shifts of its spectral lines associated with the component of its orbital velocity along the line of sight to the star.

However, YSOs are poor candidates for high-precision spectroscopic searches because of their irregular variability, rapid rotation, and faintness. The astrometric method is relatively insensitive to these characteristics, and so can be used to search for protoplanets orbiting optically visible YSOs such as the T Tauri stars. T Tauri stars are thought to be in the process of planet formation because of their youth (with ages of ~ 0.1 –10 Myr, based on their luminosities and effective temperatures), and because of the frequent presence of circumstellar disks sufficiently massive (typically 0.01–0.1 solar masses) to account for the formation of our Solar System's planets¹⁷.

The gravitational-instability mechanism begins with the formation of trailing spiral arms, which grow in amplitude until discrete clumps are formed, termed giant gaseous protoplanets (GGPPs). The wobble induced in the central protostar throughout this process has now been calculated with a three-dimensional hydrodynamics code¹⁸. The numerical model follows the nonlinear growth of non-axisymmetric (with respect to the rotation axis) density perturbations in a $0.13 M_{\odot}$ disk extending from 1 to 10 AU around a solar-mass protostar. The initial temperature profile is based on two-dimensional, radiative hydrodynamical models of disks undergoing mass accretion at astronomically inferred rates¹⁹. In the present calculations, the evolution is assumed to be locally isothermal, though the instability occurs in a very similar fashion when the disk is locally adiabatic⁸. While the inner disk is hot enough (midplane

temperature $T_m > 1,000$ K) to be gravitationally stable, the outer disk is cool enough (T_m falling to 150 K at 10 AU) for the disk to be marginally gravitationally unstable; the Toomre Q stability parameter²⁰ has a minimum value of ~ 1.2 in the outer disk.

Perturbations to the initial disk density are of the form $\delta\rho = \rho_0 a_m \cos(m\phi)$, where ρ_0 is the unperturbed density, $m = 1, 2, 3, \dots$ is a mode of amplitude a_m , ϕ is the azimuthal angle. The model was given an $m = 1$ perturbation of amplitude $a_1 = 0.01$, as well as random cell-to-cell noise at the level of $a_{m \neq 1} \approx 0.001$. The model was calculated on a spherical coordinate grid with 51 radial grid points, effectively 45 latitudinal grid points, and 64 azimuthal grid points. The outer boundary conditions (at 10 AU) were taken to absorb radial momentum at fixed density. Modes up to $m = 16$ were retained in the spherical-harmonic solution of the Poisson

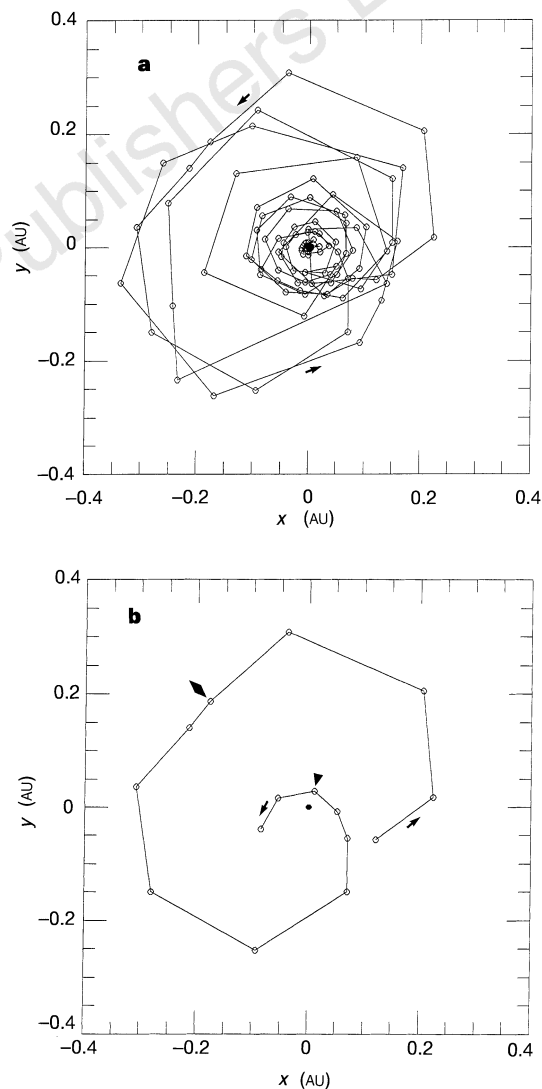


Figure 1 Positional (x - y) wobble induced in a solar-mass protostar surrounded by a protoplanetary disk undergoing a gravitational instability leading to the formation of two giant gaseous protoplanets (GGPPs). **a**, Entire evolution, up to 754 yr. Arrows denote direction of time. **b**, Final segment, from 741 to 754 yr. Filled triangle denotes when the GGPPs are on opposite sides of the protostar; filled diamond shows when they are on the same side. Central black dot is location of the centre of mass of the star and disk system. Because of the formation of multiple-Jupiter-mass GGPPs at distances of ~ 5 and ~ 10 AU, and the strong non-axisymmetry of the disk, the maximum amplitude of the anticlockwise stellar wobble becomes quite large, of the order of 0.3 AU. Such a large wobble should be detectable in optically visible YSOs at the distance (~ 140 pc) of nearby star-forming regions.

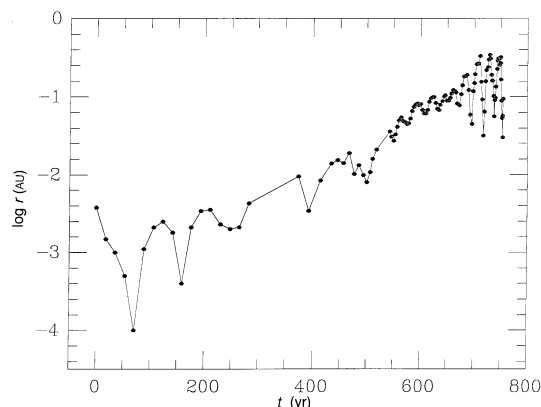


Figure 2 Time evolution of the protostellar wobble, given as the radial displacement r from the centre of mass of the system. The protostellar wobble grows roughly exponentially in time as the spiral arms grow in amplitude. After ~ 500 yr, two GGPPs form, and the amplitude of the wobble is modulated by beating between the orbital frequencies of the two GGPPs. A second model, which differed only in having additional $m = 2, 3$ and 4 initial density perturbations, behaved very similarly. (See text for details of m .)

equation for the disk's gravitational potential.

The growth of $m = 1$ modes typically results in a shift in the centre of mass of the disk away from its geometric centre. The central protostar then responds in this model by shifting in the opposite direction by an amount sufficient to preserve the exact location of the centre of mass of the system. The shifting position of the central protostar then feeds back to the disk through the star's gravitational acceleration.

Figure 1 shows the time evolution of the location of the protostar with respect to the centre of mass of the system. The protostar's wobble increases dramatically as the spiral arms grow in amplitude and form two GGPPs with masses of $\sim 0.01 M_{\odot}$, orbiting at ~ 5 AU and ~ 10 AU. The protostar's displacement from the centre of mass is minimized when the GGPPs are aligned on opposite sides of the protostar and maximized when they are on the same side (Fig. 1b). The non-axisymmetric distribution of the remainder of the disk also contributes significantly to the total protostellar wobble.

Figure 2 shows that following an initial phase of approximately exponential growth of the $m = 1$ perturbation, the mode saturates at a level that results in a protostellar wobble with an amplitude of ~ 0.1 AU. This wobble is modulated by beating between the orbital frequencies of the two GGPPs (with periods of ~ 9 yr and ~ 18 yr), yielding a period of ~ 18 yr.

The masses of the giant planets eventually formed by these $\sim 0.01 M_{\odot}$ GGPPs are uncertain, as considerable evolution remains before reaching planetary densities. In addition, future disk evolution models may well show that GGPPs can form in less-massive disks, leading to lower-mass GGPPs and smaller protostellar wobbles. Nevertheless, the model suggests that GGPP formation can lead to astrometric wobbles of the order of 0.1 AU in solar-mass YSOs. At a distance of 140 pc (for example, the Taurus star-forming region), this corresponds to an angular displacement of ~ 0.7 milliarcsec (0.7 mas), measurable with the precision of 0.1 mas per night attained by G. Gatewood's astrometric photometer at the Keck Observatory²¹.

The growth of spiral arms before GGPP formation would produce a signal about 10 times smaller (Fig. 2), ~ 70 microarcsec (70 μ as). This signal would be readily detectable by the Keck Interferometer²² or by the Space Interferometry Mission²³, which are expected to have astrometric accuracies close to ~ 10 μ as and ~ 1 μ as, respectively. However, detecting the growth of spiral arms would require fortuitous timing, because of the brief duration of

this phase, requiring the observation of an even larger ensemble of young stars.

Searching for evidence of GGPP formation around YSOs would provide a critical test of our understanding of giant-planet formation, and should clarify the relationship between giant planets and circumstellar disks, as well as any brown-dwarf protostars that would also be discovered by such a search. □

Received 24 November 1997; accepted 3 March 1998.

- Gatewood, G. Lalande 21185. *Bull. Am. Astron. Soc.* **28**, 885 (1996).
- Mayor, M. & Queloz, D. A Jupiter-mass companion to a solar-type star. *Nature* **378**, 355–359 (1995).
- Butler, R. P. & Marcy, G. W. A planet orbiting 47 Ursae Majoris. *Astrophys. J.* **464**, L153–L156 (1996).
- Butler, R. P., Marcy, G. W., Williams, E., Hauser, H. & Shirts, P. Three new “51 Pegasi-Type” planets. *Astrophys. J.* **474**, L115–L118 (1997).
- Cochran, W. D., Hatzes, A. P., Butler, R. P. & Marcy, G. W. The discovery of a planetary companion to 16 Cygni B. *Astrophys. J.* **483**, 457–463 (1997).
- Noyes, R. W. et al. A planet orbiting the star Rho Coronae Borealis. *Astrophys. J.* **483**, L111–L114 (1997).
- Kuiper, G. P. On the origin of the solar system. *Proc. Natl Acad. Sci. USA* **37**, 1–14 (1951).
- Boss, A. P. Giant planet formation by gravitational instability. *Science* **276**, 1836–1839 (1997).
- Pollack, J. B. Origin and history of the outer planets: theoretical models and observational constraints. *Annu. Rev. Astron. Astrophys.* **22**, 389–424 (1984).
- Lissauer, J. J. Timescales for planetary accretion and the structure of the protoplanetary disk. *Icarus* **69**, 249–265 (1987).
- Wetherill, G. W. Formation of the Earth. *Annu. Rev. Earth Planet. Sci.* **18**, 205–256 (1990).
- Pollack, J. B. et al. Formation of the giant planets by concurrent accretion of solids and gas. *Icarus* **124**, 62–85 (1996).
- Adams, F. C., Ruden, S. P. & Shu, F. H. Eccentric gravitational instabilities in nearly Keplerian disks. *Astrophys. J.* **347**, 959–975 (1989).
- Laughlin, G. & Bodenheimer, P. Nonaxisymmetric evolution in protostellar disks. *Astrophys. J.* **436**, 335–354 (1994).
- Trilling, D. et al. Orbital evolution and migration of giant planets: modeling extrasolar planets. *Astrophys. J.* (in the press).
- Gatewood, G. On the astrometric detection of neighboring planetary systems. *Icarus* **27**, 1–12 (1976).
- Beckwith, S. V. W., Sargent, A. L., Chini, R. S. & Güsten, R. A survey for circumstellar disks around young stellar objects. *Astron. J.* **99**, 924–945 (1990).
- Boss, A. P. & Myhill, E. A. Protostellar hydrodynamics: constructing and testing a spatially and temporally second-order accurate method. I. Spherical coordinates. *Astrophys. J. Suppl. Ser.* **83**, 311–327 (1992).
- Boss, A. P. Evolution of the solar nebula. III. Protoplanetary disks undergoing mass accretion. *Astrophys. J.* **469**, 906–920 (1996).
- Toomre, A. On the gravitational stability of a disk of stars. *Astrophys. J.* **139**, 1217–1238 (1964).
- Snyder Hale, A., Gatewood, G. D., Hale, D. D. S., Persinger, W. T. & McMillan, R. S. The multichannel astrometric photometer with spectrograph: A new instrument for the characterization of extrasolar planetary systems. *Lunar Planet. Sci. Conf. XXVIII*, 1359–1360 (1997).
- Swanson, P. *Keck Interferometer Implementation Plan* (Jet Propulsion Lab., Pasadena, 1997).
- Allen, R. J., Peterson, D. & Shao, M. *Proc. SPIE* **2871**, 504–515 (1997).

Acknowledgements. I thank D. Spergel for prompting this investigation through a cogent question, and J. Graham for improvements to the manuscript. The calculations were performed on the DEC Alpha workstations of the Carnegie Institution of Washington. This research was partially supported by NASA's Planetary Geology and Geophysics Program.

Correspondence and requests for materials should be addressed to the author (e-mail: boss@dtm.ciw.edu).

Experimental realization of a quantum algorithm

Isaac L. Chuang^{*}, Lieven M. K. Vandersypen[†], Xinlan Zhou[‡], Debbie W. Leung[§] & Seth Lloyd[§]

^{*} IBM Almaden Research Center, San Jose, California 95120, USA

[†] Solid State and Photonics Laboratory, Stanford University, Stanford, California 94305, USA

[‡] Edward L. Ginzton Laboratory, Stanford University, Stanford, California 94305, USA

[§] MIT Department of Mechanical Engineering, Cambridge, Massachusetts 02139, USA

Quantum computers^{1–5} can in principle exploit quantum-mechanical effects to perform computations (such as factoring large numbers or searching an unsorted database) more rapidly than classical computers^{1,2,6–8}. But noise, loss of coherence, and manufacturing problems make constructing large-scale quantum computers difficult^{9–13}. Although ion traps and optical cavities offer promising experimental approaches^{14,15}, no quantum algorithm has yet been implemented with these systems. Here we

report the experimental realization of a quantum algorithm using a bulk nuclear magnetic resonance technique^{16–18}, in which the nuclear spins act as ‘quantum bits’¹⁹. The nuclear spins are particularly suited to this role because of their natural isolation from the environment. Our simple quantum computer solves a purely mathematical problem in fewer steps than is possible classically, requiring fewer ‘function calls’ than a classical computer to determine the global properties of an unknown function.

We implemented the simplest possible version of the Deutsch–Jozsa (D-J) quantum algorithm⁶, which determines whether an unknown function is constant or balanced. A constant function $f(x)$ that transforms N bits of information to one bit either has output $f(x) = 0$ for all x , or $f(x) = 1$ for all x . A balanced function has $f(x) = 0$ for exactly half of its inputs, and $f(x) = 1$ for the remaining inputs. To determine with certainty whether a function is constant or balanced on a deterministic classical computer requires up to $2^{N-1} + 1$ function calls: even if half of the inputs have been examined and $f(x) = 0$ has been found for each, it cannot be concluded with certainty that the function is constant. In contrast, the D-J algorithm, as improved by Cleve *et al.*²⁰ and by A. Tapp (personal communication), allows a quantum computer to determine whether $f(x)$ is constant or balanced using only one function call.

The D-J algorithm is well illustrated by its simplest possible case, when f is a function from one bit to one bit; this is the version that we have realized (it is also the simplest instance of Simon’s algorithm⁷). There are four possible functions f , two of which are constant, $f_1(x) = 0$, $f_2(x) = 1$, and two of which have an equal number of 0 and 1 outputs: $f_3(x) = x$, $f_4(x) = \text{NOT } x$. To determine whether such a function is constant or balanced is analogous to determining whether a coin is fair (with a head on one side and a tail on the other) or fake (heads on both sides). Classically, the coin is examined twice, first on one side then on the other, to determine whether it is fair or fake. The D-J algorithm exploits quantum coherence to determine whether a quantum ‘coin’ is fair or fake while looking at it only once. The algorithm requires one ‘input’ spin and one ‘work’ spin, and is schematically represented by the quantum circuit shown in Fig. 1.

Experimentally, this quantum algorithm was implemented using the nuclear spins of the ¹H and ¹³C atoms in a carbon-13 labelled chloroform molecule (CHCl₃) as the input and work quantum bits (‘qubits’). $|0\rangle$ ($|1\rangle$) describes the spin state aligned with (against)

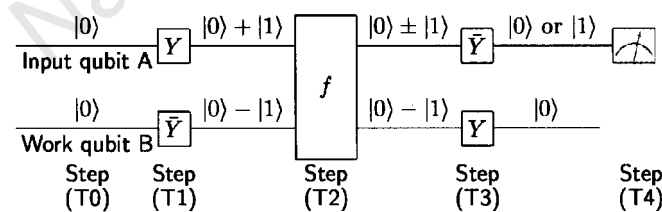


Figure 1 Quantum circuit for performing the D-J algorithm. T0, start with both the ‘input’ and ‘work’ qubits (A and B) in the state $|0\rangle$; T1, perform the transformation $Y : |0\rangle \rightarrow (|0\rangle + |1\rangle)/\sqrt{2}$, $|1\rangle \rightarrow (-|0\rangle + |1\rangle)/\sqrt{2}$, to A, and the inverse transformation $\bar{Y}$ to B, resulting in the state $\frac{1}{2}\sum_{x=0}^1 |x\rangle (|0\rangle - |1\rangle)$. The input qubit in some quantum sense registers both 0 and 1 at once. T2, call the function: apply f to A, and add the result to B modulo 2. As long as the quantum logic operations needed to evaluate f are carried out coherently, the work qubit now contains in some quantum sense the outputs of f on all possible inputs, an effect that Deutsch termed ‘quantum parallelism’. The two qubits are now in the state $\frac{1}{2}\sum_{x=0}^1 |x\rangle (|0 + f(x)\rangle - |1 + f(x)\rangle) = \frac{1}{2}\sum_{x=0}^1 (-1)^{f(x)} |x\rangle (|0\rangle - |1\rangle)$. T3, perform the inverse of the transformations of (T1), thereby taking the qubits out of their superposition states. If f is constant, then the factors $(-1)^{f(x)}$ are either all +1 or all -1, and the result of the transformation in this step is the state $\pm |00\rangle$. If f is balanced, then exactly half of the factors $(-1)^{f(x)}$ are +1 and half are -1, and the result of the transformation is the state $\pm |10\rangle$. T4, read out A. If it is 0, then f is constant; if it is 1, then f is balanced.

an externally applied, strong static magnetic field $\mathbf{B}_0$ in the $+\hat{z}$ direction. The reduced hamiltonian for this 2-spin system is to an excellent approximation given by $(\hbar=1)^{21}$: $\hat{\mathcal{H}} = -\omega_A \hat{I}_{zA} - \omega_B \hat{I}_{zB} + 2\pi J \hat{I}_{zA} \hat{I}_{zB} + \hat{\mathcal{H}}_{\text{env}}$. The first two terms describe the free precession of spin A (¹H) and B (¹³C) about $-\mathbf{B}_0$ with frequencies $\omega_A/2\pi \approx 500$ MHz and $\omega_B/2\pi \approx 125$ MHz. $\hat{I}_{zA}$ is the angular momentum operator in the $+\hat{z}$ direction for A. The third term describes a scalar spin–spin coupling of the two spins of $J \approx 215$ Hz. $\hat{\mathcal{H}}_{\text{env}}$ represents coupling to the environment, including interactions with the chlorine nuclei, and also higher order terms in the spin–spin coupling, which can be disregarded (as will be described below).

The five theoretical steps of the quantum algorithm, (T0)–(T4), were experimentally implemented as follows.

E0. An initial state is prepared with a 200 mM, 0.5 ml sample of chloroform dissolved in *d*6-acetone, at room temperature and standard pressure. The $O(10^{18})$ molecules in this bulk sample can be thought of as being independent single quantum computers, all functioning simultaneously. The theoretically ideal result is obtained when the spins in all the molecules start out in the 00 state—that is, all spins aligned in the $+\hat{z}$ direction. Because the experiment is performed at room temperature, however, the initial density matrix ρ for the thermally equilibrated system has populations $\text{diag}(\rho) = [n_{00}, n_{01}, n_{10}, n_{11}]$ in the 00, 01, 10 and 11 states, respectively, where ρ is the density matrix, and n_i are proportional to $e^{-E_i/kT}/2^N \approx (1 - E_i/kT)/2^N$, with E_i the energy of state i , and $N = 2$ the number of qubits used in our experiment. A variety of techniques exist to extract from this thermal state just the signal from the molecules in the 00 state^{16,17}; we adopted the method of ‘temporal averaging’²², which involves the summation of three experiments in which the populations of the 01, 10 and 11 states are cyclically permuted before performing the computation. The essential observation is that $[n_{00}, n_{01}, n_{10}, n_{11}] + [n_{00}, n_{11}, n_{01}, n_{10}] + [n_{00}, n_{10}, n_{11}, n_{01}] = \alpha[1, 1, 1, 1] + \delta[1, 0, 0, 0]$, where $\alpha = n_{01} + n_{10} + n_{11}$ is a background signal which is not detected, and $\delta = 3n_{00} - \alpha$ is a deviation from the uniform background whose signal behaves effectively like the desired pure quantum state, $|00\rangle$. The permutations are performed using methods similar to those used for the computation, described next. This technique avoids the technical difficulties of detecting the signal from a single nuclear spin, and allows a sample at room temperature, which produces an easily detectable signal, to be used for quantum computation.

Note that although this method requires $f(x)$ to be evaluated three times, it is actually not necessary. Although step (T0) stipulates a pure initial state $|00\rangle$, the algorithm works equally well if the input qubit is initially $|1\rangle$; furthermore, when the work qubit is initially $|1\rangle$, it fails, and cannot distinguish constant from balanced functions, but this does not interfere with other computers which have worked. Thus, a thermal state is a good initial state for this algorithm, and only one experiment needs to be done. Data from both thermal and pure state inputs are presented below.

E1. Pulsed radiofrequency electromagnetic fields are applied to transform the qubits as prescribed in (T1). These fields, oriented in the $\hat{x} - \hat{y}$ plane perpendicular to $\mathbf{B}_0$, selectively address either A or B by oscillating at frequency ω_A or ω_B . Classically, a radiofrequency pulse along $\hat{y}$ (for example) rotates a spin about that axis by an angle proportional to $\sim tP$, the product of the pulse duration t and pulse power P . In the ‘bar magnet’ picture, a $\pi/2$ pulse along $\hat{y}$ (we shall call this Y) causes a $\hat{z}$ oriented spin to be rotated by 90° , onto $\hat{x}$ (similarly, we shall let $\bar{Y}$ denote $\pi/2$ rotations about $-\hat{y}$, and X denote $\pi/2$ rotations about $\hat{x}$, and so forth; subscripts will identify which spin the operation acts upon). This description of the state is classical in the sense that a bar magnet always has a definite direction. In reality, however, a nuclear spin is a quantum object, and instead of being aligned along $\hat{x}$, it is actually in a superposition of being up and down, $(|0\rangle + |1\rangle)/\sqrt{2}$. Likewise, a spin classically

described as being along $-\hat{x}$ is actually in the state $(|0\rangle - |1\rangle)/\sqrt{2}$. (E1) thus consists of applying the two radiofrequency pulses $Y_A Y_B$. E2. The function $y \rightarrow y \oplus f(x)$ is implemented using radiofrequency pulses and spin-spin interaction. Recall that spin A represents the input qubit x , and B the work qubit y where f stores its output. f_1 is implemented as $\tau/2 - X_B X_B - \tau/2 - X_B X_B$, to be read from left to right, where $\tau/2$ represents a time interval of $1/4f \approx 1.163$ ms, during which coupled spin evolution occurs. Dashes are for readability only, and typical pulse lengths were 10–15 μ s. This is a well known refocusing²³ pulse sequence which performs the identity operation. f_2 is $\tau/2 - X_B X_B - \tau/2$, similar to f_1 but without the final pulses, so that B is inverted. f_3 is $Y_B - \tau - Y_B X_B - Y_A X_A Y_A$, which implements a ‘controlled-NOT’ operation, in which B is inverted if, and only if, A is in the $|1\rangle$ state. The naive ‘bar magnet’ picture can be used to get a feeling for how this works in case the inputs are 00 or 10, for which the subsequence $Y_B - \tau - X_B$ suffices (note that after (E1), both spins are not just $|0\rangle$ or $|1\rangle$ but in a superposition of both, in which case the extra pulses of f_3 are necessary¹⁶). First, Y_B rotates B to $+\hat{x}$. B then precesses in the $\hat{x} - \hat{y}$ plane, about $-\hat{z}$. Owing to the spin-spin coupling, B precesses slightly slower (faster) if $A = 0$ ($A = 1$). After τ seconds, B reaches $+\hat{y}$ ($-\hat{y}$) in the rotating frame. X_B then rotates B to $+\hat{z}$ ($-\hat{z}$), that is, to 0 or 1, where the final state of B depends on the input A . The precise quantum description is easily obtained by multiplying out the

unitary rotation matrices. Finally, f_4 is implemented as $Y_B - \tau - Y_B X_B - Y_A X_A Y_A$, which is similar to f_3 but leaves B inverted.

E3. The inverse of (E1) is done by applying the radiofrequency pulses $Y_A Y_B$ to take both spins back to $\pm\hat{z}$. Spin A , which was $|0\rangle$ at the input, is thus transformed into $|0\rangle$ or $|1\rangle$ for constant or balanced functions respectively.

E4. The result is read out by applying a read-out pulse X_A to bring spin A back into the $\hat{x} - \hat{y}$ plane. The time varying voltage $V(t)$ induced by the precession of spin A about $-\mathbf{B}_0$ is recorded by a phase sensitive pick-up coil. Inspection of the spectrum of $V(t)$ after a single experiment run and an appropriate read-out pulse, immediately reveals whether $f(x)$ is constant or balanced (Fig. 2).

We also characterized the entire deviation density matrix $\rho_\Delta \equiv \rho - \text{Tr}(\rho)I/4$ (Fig. 3) describing the final 2-qubit state. These results unambiguously demonstrate the complete proper functioning of the quantum algorithm, and provide data for the error analysis described below.

Quantum computation requires that a coherent superposition be preserved for the duration of the computation. This requires a highly isolated quantum system (small $\mathcal{H}_{\text{env}}$), and fortunately, nuclear spins are naturally well isolated from their environment. Phase randomization due to $\mathbf{B}_0$ inhomogeneities was minimized by using about 30 electromagnetic coils to shim the static field to be constant to about one part in 10^9 over the sample volume. The

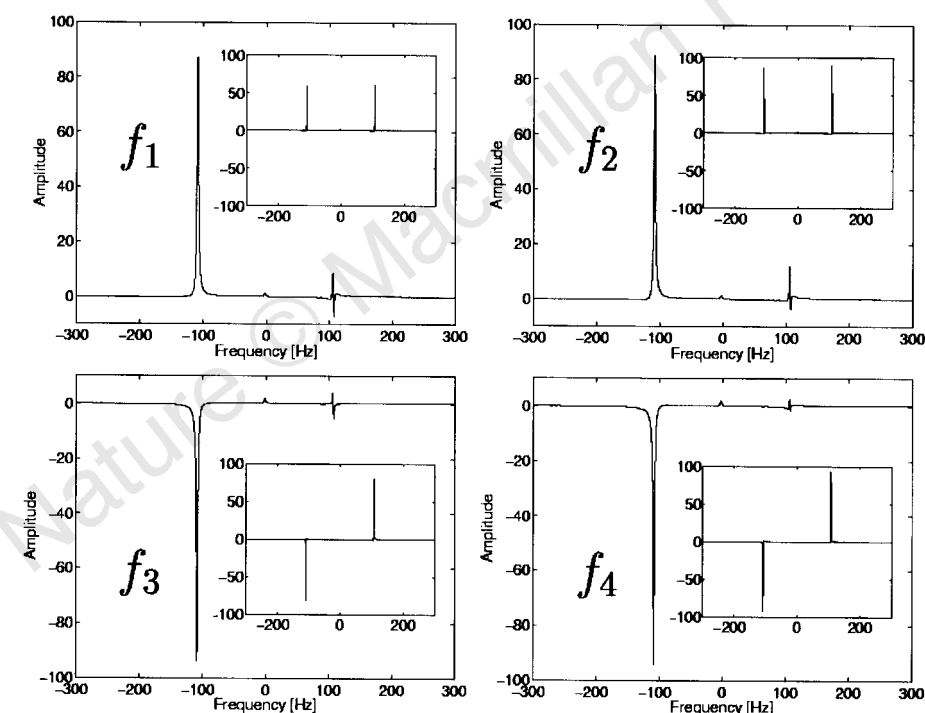


Figure 2 Proton spectrum after completion of the D-J algorithm and a single read-out pulse X_A , with an effectively pure initial state $|00\rangle$ and with a thermal initial state (inset). The low/high-frequency lines correspond to the transitions $|00\rangle \leftrightarrow |10\rangle$ ($|01\rangle \leftrightarrow |11\rangle$). The frequency is relative to 499,755,169 Hz, and the amplitude has arbitrary units. The spectrum is the Fourier-transformed time varying voltage $V(t)$, induced in the pick up coil by the precession of spin A about $-\mathbf{B}_0$, at frequency ω_A , after the read-out pulse X_A . $V(t)$ is given by $V(t) \approx V_0 \text{Tr}[e^{-i\omega_A t} e^{-i(\pi/2)\hat{S}_x} \rho(0) e^{i(\pi/2)\hat{S}_x} e^{i\omega_A t} \times (-i\sigma_{xA} - \sigma_{yA})]$, where $\sigma_{\{x,y\}}$ are Pauli matrices, and $\rho(0)$ is the density matrix of the state immediately before the readout pulse. By this convention, a spectral line for spin A is real and positive (negative) when spin A is $|0\rangle$ ($|1\rangle$) right before the X_A read-out pulse. Experiments were performed at Stanford University using an 11.7-Tesla Oxford Instruments magnet and a Varian^{UNIVERSITY} Inova spectrometer with a triple-resonance probe. ^{13}C -labelled CHCl_3 was obtained from Cambridge Isotope Laboratories.

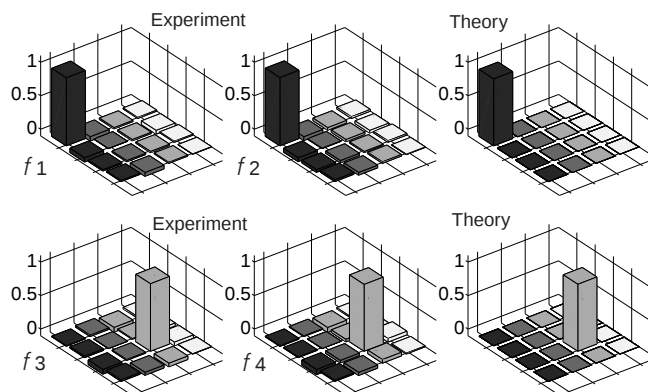


Figure 3 Experimentally measured and theoretically expected deviation density matrices after completion of the D-J algorithm. The diagonal elements represent the normalized populations of the states $|00\rangle$, $|01\rangle$, $|10\rangle$ and $|11\rangle$ (from left to right). The off-diagonal elements represent coherences between different states. The magnitudes are shown with the sign of the real component; all imaginary components were small. The deviation density matrix was obtained from the integrals of the proton and carbon spectral lines, acquired for a series of 9 experiments with different read-out pulses for each spin (quantum state tomography²⁴). The observed experimental non-idealities can be quantified as follows. In the experiments, the normalized pure-state population (ideally equal to 1), varied from 0.998 to 1.019. The other deviation density matrix elements (ideally 0), were smaller than 0.075 in magnitude. The relative error ϵ on the experimental pure-state output density matrix ρ_{exp} , defined as $\epsilon = \|\rho_{\text{exp}} - \rho_{\text{theory}}\|/\|\rho_{\text{theory}}\|$, varied between 8 and 12%.

longitudinal and transverse relaxation time constants T_1 and T_2 were measured using standard inversion–recovery and Carr–Purcell–Meiboom–Gill pulse sequences²³, giving $T_1 \approx 19$ and 25 s, and $T_2 \approx 7$ and 0.3 s, respectively, for proton and carbon; these were much longer than required for our experiment, which finished in about 7 ms.

The single most important source of errors in the experiments was the radiofrequency field inhomogeneity and pulse-length calibration imperfections. A direct measure of this inhomogeneity is the ~ 200 - μ s time constant of the exponentially decaying envelope observed from applying a single pulse as a function of pulse length. Including the population permutation sequence, about 7 pulses are applied to each nucleus, with a cumulative duration of ~ 70 – 100 μ s.

The second most important contribution to errors is the low carbon signal-to-noise ratio, signal peak height/r.m.s. noise ≈ 35 , versus about 4,300 for proton. The carbon signal was much weaker because the carbon gyromagnetic ratio is 4 times smaller, and the carbon receiver coil is mounted more remotely from the sample. Smaller contributions to errors came from incomplete relaxation between subsequent experiments, carrier frequency offsets and numerical errors in the data analysis.

For this small-scale quantum computer, imperfections were dominated by technology, rather than by fundamental issues. However, NMR quantum computers larger than about 10 qubits will require creative new approaches, because the signal strength decays exponentially with the number of qubits in the machine, using current schemes^{24,25}: for N spins, the signal from the initial state $00\dots 0$ is proportional to $n_{00\dots 0} \propto NZ^{-N}$, where the single spin partition function $Z \approx 2$ at high temperatures. Furthermore, coherence times typically decrease for larger molecules, whereas the average logic gate duration increases. Nevertheless, there is hope; for example, because of the ensemble nature of the NMR approach, the output result can be inferred as long as a distinguishable majority of the molecules reach the correct final state. Creating an effective pure state is thus not always necessary, as we have demonstrated. Optical pumping and other cooling techniques can also be used to prepolarize the sample to increase the output signal amplitude, because $Z \approx 1$ at low temperatures. Quantum computation poses an interesting and relevant experimental challenge for the future.

Note added in proof: During the course of this work, we became aware of a closely related experiment by J. A. Jones and M. Mosca²⁶.

Received 21 January; accepted 18 March 1998.

- Deutsch, D. Quantum theory, the Church-Turing principle and the universal quantum computer. *Proc. R. Soc. Lond. A* **400**, 97–117 (1985).
- Shor, P. Algorithms for quantum computation: discrete logarithms and factoring. *Proc. 35th Annu. Symp. on Found. of Computer Science* 124–134 (IEEE Comp. Soc. Press, Los Alamitos, CA, 1994).
- Divincenzo, D. P. Quantum computation. *Science* **270**, 255–261 (1995).
- Lloyd, S. Quantum-mechanical computers. *Sci. Am.* **273**, 44–50 (1995).
- Ekert, A. & Jozsa, R. Quantum computation and Shor's factoring algorithm. *Rev. Mod. Phys.* **68**, 733–753 (1996).
- Deutsch, D. & Jozsa, R. Rapid solution of problems by quantum computation. *Proc. R. Soc. Lond. A* **439**, 553–558 (1992).
- Simon, D. On the power of quantum computation. *Proc. 35th Annu. Symp. on Found. of Computer Science* 116–124 (IEEE Comp. Soc. Press, Los Alamitos, CA, 1994).
- Grover, L. K. Quantum computers can search arbitrarily large databases by a single query. *Phys. Rev. Lett.* **79**, 4709–4712 (1997).
- Uhrig, W. G. Maintaining coherence in quantum computers. *Phys. Rev. A* **51**, 2, 992–997 (1995).
- Chuang, I. L., Laflamme, R., Shor, P. & Zurek, W. H. Quantum computers, factoring, and decoherence. *Science* **270**, 1633–1635 (1995).
- Landauer, R. Dissipation and noise immunity in computation and communication. *Nature* **335**, 779–784 (1988).
- Landauer, R. Is quantum mechanics useful? *Phil. Trans. R. Soc. Lond. A* **335**, 367–376 (1995).
- Palma, G. M., Suominen, K.-A. & Ekert, A. K. Quantum computers and dissipation. *Proc. R. Soc. Lond. A* **452**, 567–584 (1996).
- Monroe, C., Meekhof, D. M., King, B. E., Itano, W. M. & Wineland, D. J. Demonstration of a fundamental quantum logic gate. *Phys. Rev. Lett.* **75**, 4714–4717 (1995).
- Turchette, Q. A., Hood, C. J., Lange, W., Mabuchi, H. & Kimble, H. J. Measurement of conditional phase shifts for quantum logic. *Phys. Rev. Lett.* **75**, 4710–4713 (1995).
- Gershenfeld, N. & Chuang, I. L. Bulk spin-resonance quantum computation. *Science* **275**, 350–356 (1997).
- Cory, D. G., Price, M. D., Fahmy, A. F. & Havel, T. F. Nuclear magnetic resonance spectroscopy: an experimentally accessible paradigm for quantum computing. *Physica D* (in the press; LANL E-print quant-ph/9709001.gov, 1997).

- Cory, D. G., Fahmy, A. F. & Havel, T. F. Ensemble quantum computing by NMR spectroscopy. *Proc. Natl Acad. Sci. USA* **94**, 1634–1639 (1997).
- Lloyd, S. A potentially realizable quantum computer. *Science* **261**, 1569–1571 (1993).
- Cleve, R., Ekert, A., Macchiavello, C. & Mosca, M. *Proc. R. Soc. Lond. A* **454**, 339–354 (1998). LANL E-print quant-ph/9708016.
- Slichter, C. P. *Principles of Magnetic Resonance* (Springer, Berlin, 1990).
- Knill, E., Chuang, I. L. & Laflamme, R. Effective pure states for bulk quantum computation. *Phys. Rev. A* **57**(5), May (1998).
- Ernst, R. R., Bodenhausen, G. & Wokaun, A. *Principles of Nuclear Magnetic Resonance in One and Two Dimensions* (Oxford Univ. Press, Oxford, 1994).
- Chuang, I. L., Gershenfeld, N., Kubinec, M. G. & Leung, D. W. Bulk quantum computation with nuclear magnetic resonance: Theory and experiment. *Proc. R. Soc. Lond. A* **454**, 447–467 (1998).
- Warren, W. S. The usefulness of NMR quantum computing. *Science* **277**, 1688–1690 (1997).
- Jones, T. F. & Mosca, M. Implementation of a quantum algorithm to solve Deutsch's problem on a nuclear magnetic resonance quantum computer. *J. Chem. Phys.* (in the press; LANL E-print quant-ph/9801027).

Acknowledgements. We thank A. Pines and M. Kubinec for discussion. This work was supported by DARPA under the NMRQC and QUIC initiatives. L.V. gratefully acknowledges a Francqui Fellowship of the Belgian American Educational Foundation and a Yansouni Family Fellowship.

Correspondence and requests for materials should be addressed to I.L. (e-mail: ichuang@almaden.ibm.com).

Spontaneous formation of ordered structures in thin films of metals supported on an elastomeric polymer

Ned Bowden*, Scott Brittain*, Anthony G. Evans†, John W. Hutchinson† & George M. Whitesides*

* Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138, USA

† Department of Engineering and Applied Sciences, Harvard University, Pierce Hall, Cambridge, Massachusetts 02138, USA

Spontaneous generation of complex order in apparently simple systems is both arresting and potentially useful^{1–11}. Here we describe the appearance of complex, ordered structures induced by the buckling of thin metal films owing to thermal contraction of an underlying substrate. We deposit the films from the vapour phase on a thermally expanded polymer (polydimethylsiloxane, PDMS). Subsequent cooling of the polymer creates compressive stress in the metal film that is relieved by buckling with a uniform wavelength of 20–50 micrometres. The waves can be controlled and orientated by relief structures in the surface of the polymer, which can set up intricate, ordered patterns over large areas. We can account qualitatively for the size and form of the patterned features in terms of the non-uniform stresses developed in the film near steps on the polymer substrate. This patterning process may find applications in optical devices such as diffraction gratings and optical sensors, and as the basis for methods of strain analysis in materials.

Thin metal films—typically 50-nm-thick layers of gold with a 5-nm adhesion interlayer of titanium or chromium—were deposited onto PDMS by electron beam evaporation (Fig. 1). The metal source heats and expands the PDMS substrate before and during deposition. We believe that local heating of the surface of the PDMS also slightly modifies its mechanical properties; perhaps by introducing new crosslinks¹². After cooling to ambient temperature, the surface appeared frosted, because of light scattering from a network of periodic surface waves (Fig. 2a)¹³. Similar waves were found with a variety of metals, including nickel, aluminum, titanium and chromium. The waves almost disappeared when the sample was reheated to 110 °C, but reformed on cooling. They also formed on PDMS externally heated to 300 °C during the evaporation of metal. Conversely, when the PDMS was cooled to 0 °C during evaporation of metal, waves did not arise; this observation verifies the central importance of the thermal excur-

sion. The waves had periodicities between 20 and 50 μm ; their depths, measured from the crest to the trough, ranged from 1.5 μm for evaporations performed without external heating, to 3.9 μm for evaporations conducted at 300 $^{\circ}\text{C}$. We propose that these waves result from redistribution, by buckling, of compressive stresses that develop in the surface of the sample on cooling from the evaporator temperature to ambient temperature. These waves are similar to those formed in the wrinkling of skin sheets in sandwich structures¹⁴, but different from the waves in buckling-driven film delamination¹⁵ because the metal films remain attached to the PDMS.

On flat, unconstrained PDMS, the waves were disordered, except near an edge. When the PDMS was attached to a glass slide that

constrained its expansion on one face, the ordering that occurred reflected, in part, this constraint. To allow the PDMS to contract isotropically, we separated the PDMS, which had been coated with gold and which showed a pattern of waves, from the glass slide with a razor blade. The PDMS was placed on a paper tissue, heated until the waves almost disappeared, and then allowed to cool again. The PDMS expanded and contracted isotropically. The waves reformed with dimensions similar to those observed before, but, in general, in patterns ordered only near relief structures in the PDMS. Strong ordering occurred on evaporating the metal film onto PDMS having a bas-relief pattern on its surface (Fig. 2b–f); these patterns were created by casting PDMS against a 2–20 μm thick patterned photoresist layer^{16,17}. The pattern of waves made a transition from

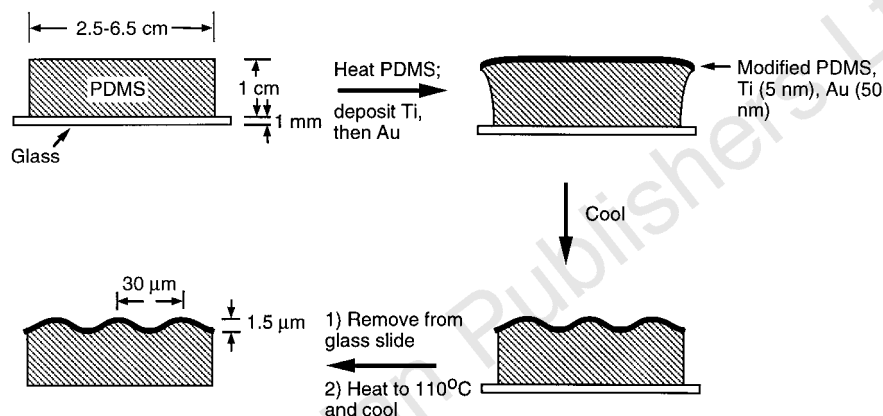


Figure 1 Preparation of metal films on PDMS. A thick ($1.0 \times 2.5 \times 6.5$ cm) layer of PDMS was allowed to adhere to a thin glass slide and was loaded into an electron beam evaporator; the chamber was evacuated to a pressure of 10^{-6} torr. The electron beam heated the metal sources; this heat also reached the PDMS and caused it to expand. Here we exaggerate the expansion of the PDMS for clarity; in fact it expands by $<1\%$. The metals were deposited on the expanded PDMS; a

typical film comprised 50 $\text{\AA}$ of titanium evaporated at $1 \text{ \AA s}^{-1}$, followed by 500 $\text{\AA}$ of gold evaporated at $3 \text{ \AA s}^{-1}$. After the evaporation, the PDMS cooled and contracted; waves formed on the surface. The PDMS was separated from the glass slide with a razor blade, heated to 110 $^{\circ}\text{C}$, and cooled to ambient temperature. The pattern of waves before and after separation, heating, and cooling were similar.

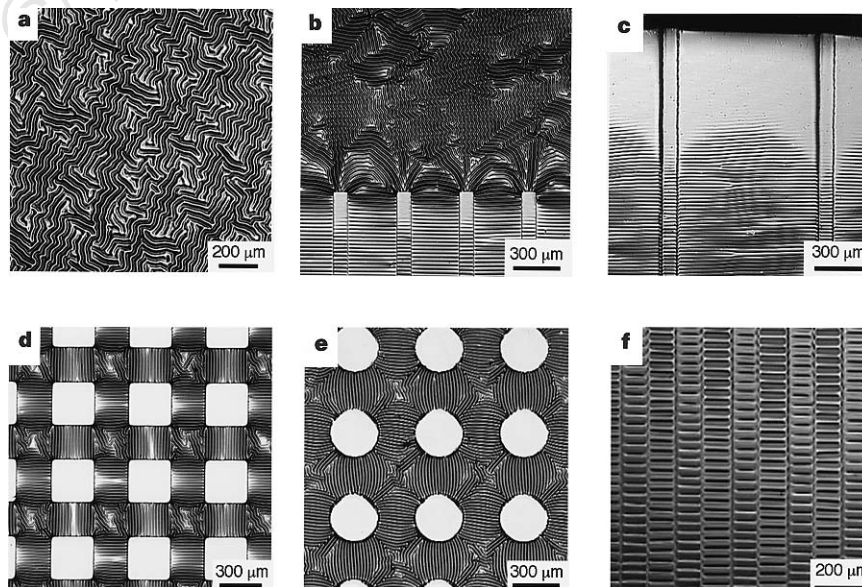


Figure 2 Optical micrographs showing representative patterns of waves that formed when the metals were evaporated onto warm ($\sim 110^{\circ}\text{C}$) PDMS, and the sample then cooled to room temperature. **a**, Disordered regions covered flat PDMS far from any steps or edges. **b**, Micrograph showing the transition from disordered waves to waves ordered by rectangular ridges (100 μm wide and 10–20 μm high; separated by 300 μm). **c**, A flat, waveless region of gold near an edge in the PDMS gradually became a system of waves ordered by the rectangular ridges (100 μm wide and 10–20 μm high; separated by 800 μm). **d**, Flat squares

(300 μm on each side) and circles (150 μm in radius) elevated by 10–20 μm relative to the surface showed no buckling on the plateaux, but ordered patterns of waves on the recessed regions between them. **f**, Rectangular ridges (100 μm wide and 10–20 μm high; separated by 100 μm) aligned the waves parallel to the direction of the raised portions of the PDMS. These pictures are representative of the patterns that form over the whole surface area for each sample (up to 25 cm^2). The PDMS was coated with 5 nm of titanium or chromium and 50 nm of gold.

disordered (Fig. 2a) to ordered (Fig. 2b–f) when located within 200–600 μm of a step or edge.

We have developed a model that relates the spatially non-uniform stresses in a metal film on patterned PDMS to the wave patterns. This model correctly estimates the circumstances in which waves initiate, as well as their orientation and wavelength. We describe this model for two systems: (1) a single thin, stiff surface film on a thick, homogeneous compliant substrate, representing Au on PDMS, and (2) a multilayer film, representing Au on PDMS having a surface layer with modified properties (Fig. 3).

If the PDMS were perfectly smooth and the film unbuckled, then at temperatures T below the deposition temperature T_D , the film would be in a state of uniform, equi-biaxial compressive stress σ_o (Pa) given by equation (1) (ref. 18):

$$\sigma_o = \frac{E_m(\alpha_p - \alpha_m)(T_D - T)}{(1 - \nu_m)} \quad (1)$$

Here the subscripts m and p refer to the metal film and PDMS, respectively, ν (unitless) is the Poisson's ratio, α ($^{\circ}\text{C}^{-1}$) is the coefficient of thermal expansion, and E (Pa) is the Young's modulus. The parameters for gold and PDMS are: $E_m = 82$ GPa, $E_p = 20$ MPa, $\nu_m = 0.33$ and $\nu_p = 0.48$. The equi-biaxial compressive film stress arises from the considerable mismatch of the coefficients of thermal expansion of the PDMS and the film ($\alpha_p \approx 20\alpha_m$).

As the temperature drops and the compressive stresses in the film increase, buckling starts where the maximum principal compressive stress attains the critical value, σ_{crit} (Pa; equation (2))¹⁴. The associated sinusoidal wave pattern, aligned perpendicular to the direction of maximum compression, has wavelength L (m) given by equation (3)¹⁴. Here t (m) is the film thickness.

$$\sigma_{\text{crit}} \approx 0.52 \left(\frac{E_m}{(1 - \nu_m^2)} \right)^{1/3} \left(\frac{E_p}{(1 - \nu_p^2)} \right)^{2/3} \quad (2)$$

$$L \approx 4.36t \left(\frac{E_m(1 - \nu_p^2)}{E_p(1 - \nu_m^2)} \right)^{1/3} \approx 4.4t \left(\frac{E_m}{E_p} \right)^{1/3} \approx 61t \quad (3)$$

Because the film is stiff relative to the PDMS, the wavelength is many times the film thickness. Equation (3) yields an estimate: $L \sim 3.4 \mu\text{m}$. As the observed value of L is approximately nine times larger than this estimate, we conclude that this model is oversimplified. In practice, we believe that the heating occurring during the metal

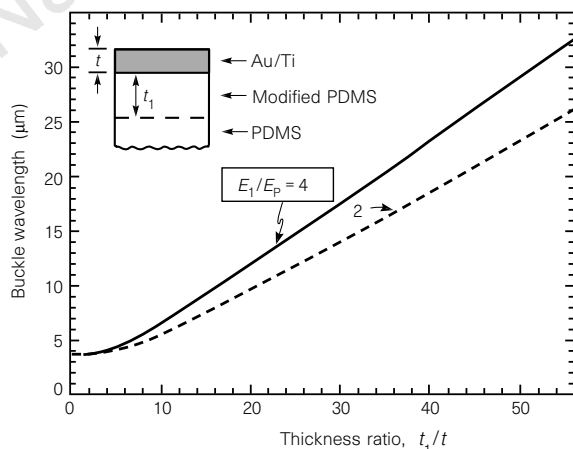


Figure 3 How the wavelength changes with a modified layer of PDMS at the surface. We assume that a uniform, modified layer of PDMS of thickness t_1 and Young's modulus E_1 exists between the gold of thickness t and PDMS. The wavelength increases as the thickness of the modified layer of PDMS is increased relative to the thickness of the gold film, and as the ratio E_1/E_p is increased.

deposition also modifies the top surface of the PDMS (plausibly by thermal crosslinking), generates compressive stress, and alters its Young's modulus, perhaps to a depth of the order of $1 \mu\text{m}$. This modification creates a multilayered film with an effective bending stiffness larger than that of the thin metal film, and increases the wavelength (Fig. 3). A modified PDMS layer of thickness $2 \mu\text{m}$ (40 times that of the Au) with a plausible Young's modulus ($E_1 \approx 80$ MPa $\approx 4E_p$) increases the wavelength by a factor of 7, in satisfactory agreement with the experimental result. Uncertainty with respect to the properties and structure of the modified PDMS layer makes it difficult to provide a more quantitative comparison.

In an equi-biaxial state of stress, there is neither a preferred orientation for the waves, nor a reason for the waves to form systematic patterns. When steps are present, however, the stress in

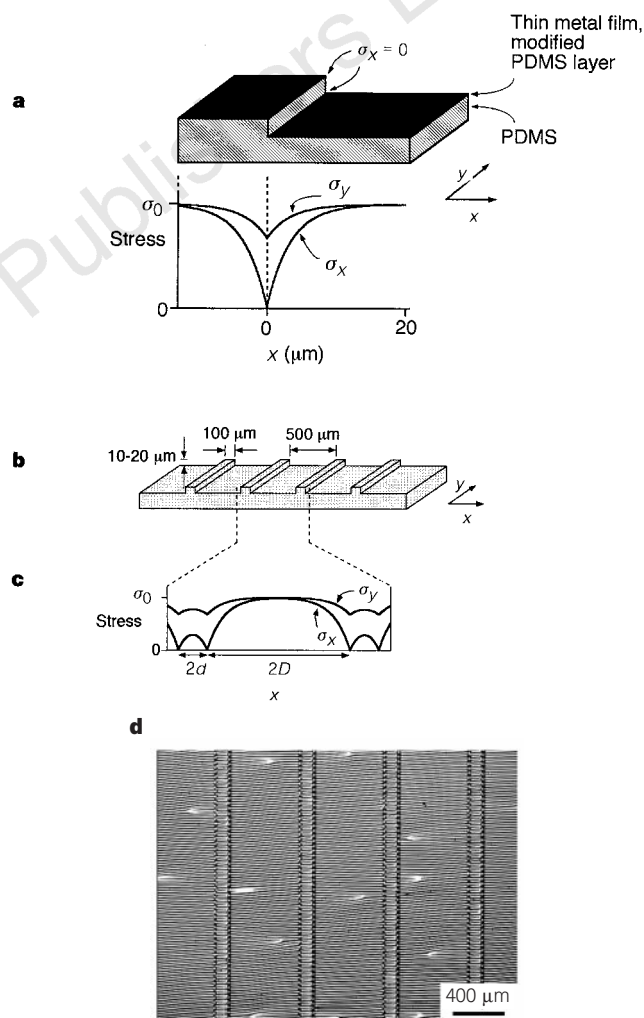


Figure 4 Ordering of waves on a surface of PDMS patterned in a bas-relief structure. **a**, We assume that in-plane displacement normal to the line of a step in the PDMS (2–20 μm high) relieves the compressive stress in the metal film; that is, $\sigma_x = 0$ at the step. The graph shows σ_x and σ_y as a function of x at a step in the PDMS (from equations (4a) and (4b)). The variation in σ_y as a function of x is due to a Poisson effect where the stress in the y -direction is partially relieved by expansion of the metal film into the x -direction. **b**, The surface of the PDMS was patterned into regions 100- μm wide that were raised by 10–20 μm and separated by 500 μm ; 5 nm of titanium and 50 nm of gold were deposited on this surface. **c**, The stress in the system followed equations (6a) and (6b). The pattern of waves in Fig. 2c follows the distribution of stress predicted by equations (6a) and (6b). **d**, An optical micrograph picture of the PDMS surface after buckling. The waves are aligned perpendicular to the steps and in the direction of the greater stress.

the film will no longer be uniform or equi-biaxial. There is a strong orientation to the stress in the vicinity of the steps, with an associated maximum principal compressive stress direction at each point. The wave pattern develops with crests aligned perpendicular to the direction of maximum compressive stress.

The non-uniform pre-buckling stress around steps is derived from a model that approximates the influence of the substrate on the thin film as an attached elastic foundation of springs exerting only tangential forces. The PDMS offers little resistance to displacement of the film in the direction perpendicular to the step; this displacement relieves the stress in that direction, becoming zero at the step itself. A solution for the stresses on either side of an infinitely long isolated step coincident with the y -axis (Fig. 4a) gives equations (4a) and (4b), where σ_x is the stress in the x -direction, σ_y is the stress in the y -direction, and x is measured starting from the step. The transition length l (m) characterizing the distribution of stress from the step to the remote smooth film is given by equation (5).

$$\sigma_x = -\sigma_0[1 - e^{-|x|/l}] \quad (4a)$$

$$\sigma_y = -\sigma_0[1 - \nu_m e^{-|x|/l}] \quad (4b)$$

$$l \approx 0.3t \left[\frac{E_m(1 - \nu_p^2)}{E_p(1 - \nu_m^2)} \right] \approx 0.3t \left[\frac{E_m}{E_p} \right] \approx 1,060t \quad (5)$$

The transition length is 17 times the buckle wavelength, L , for the gold/PDMS system.

The predictability of the wave patterns is demonstrated by analysis of an array of long straight strips parallel to the y -axis, width $2d$, separated by $2D$ (Fig. 4b). The stress distribution in the film before buckling is given by equations (6a) and (6b), for ($|x| < d$) where x is measured from the centre of the strip (Fig. 4c).

$$\sigma_x = -\sigma_0 \left[1 - \frac{\cosh(x/l)}{\cosh(d/l)} \right] \quad (6a)$$

$$\sigma_y = -\sigma_0 \left[1 - \nu_m \frac{\cosh(x/l)}{\cosh(d/l)} \right] \quad (6b)$$

The stresses between the strips are given by the same formulae, but with d replaced by D , and with x now measured from the centre of that region. We note that the compressive stress in the y -direction remains relatively large, with a minimum, $-\sigma_0(1 - \nu_m)$ at the step; the compressive stress in the x -direction is smaller everywhere, and zero at the step. Consistent with σ_y being larger than σ_x , the wave pattern shows crests aligned perpendicular to the y -axis (Fig. 4d).

The process described here provides a remarkable example of the spontaneous generation of complexity. The regularity in the waves reflects the uniformity in the physical properties and dimensions of the materials. The ability to control the orientation and periodicity of these waves by changing these parameters, and by patterning the surface of the PDMS using straightforward techniques, makes this system eminently controllable. We believe that this process offers potential to generate planar and non-planar surfaces patterned in 1–100 μm features over many square centimetres. Such patterns are interesting for their potential applications in sensors and optical components (for example, diffraction gratings): they also allow mapping of material properties in two dimensions. Most interestingly, such patterns offer the opportunity to study the generation of complex ordered structure from simple patterns. $\square$

Received 7 January; accepted 16 March 1998.

- Whitesides, G. M., Mathias, J. P. & Seto, C. T. Molecular self-assembly and nanochemistry: a chemical strategy for the synthesis of nanostructures. *Science* **254**, 1312–1319 (1991).
- Desiraju, G. R. *Crystal Engineering: The Design of Organic Solids* (Elsevier, New York, 1989).
- van Blaaderen, A., Ruel, R. & Wiltzius, P. Template-directed colloidal crystallization. *Nature* **385**, 321–324 (1997).
- Bain, C. D. *et al.* Formation of monolayer films by the spontaneous assembly of organic thiols from solution onto gold. *J. Am. Chem. Soc.* **111**, 321–335 (1989).
- Voet, D. & Voet, J. G. *Biochemistry* (Wiley, New York, 1995).

- Davis, M. E., Lobo, R. F. Zeolite and molecular sieve synthesis. *Chem. Mater.* **4**, 756–768 (1992).
- Lewis, D. W., Willock, D. J., Catlow, C. R. A., Thomas, J. M. & Hutchings, G. J. De novo design of structure-directing agents for the synthesis of microporous solids. *Nature* **382**, 604–606 (1996).
- Makse, H. A., Havlin, S., King, P. R. & Stanley, H. E. Spontaneous stratification in granular mixtures. *Nature* **386**, 379–382 (1997).
- Jaeger, H. M. & Nagel, S. R. Physics of the granular state. *Science* **255**, 1523–1531 (1992).
- Zik, O., Levine, D., Shtrikman, S. G. & Stavans, J. Rotationally induced segregation of granular materials. *Phys. Rev. Lett.* **73**, 644–647 (1994).
- Knight, J. B., Jaeger, H. M. & Nagel, S. R. Vibration-induced size separation in granular media: the convection connection. *Phys. Rev. Lett.* **70**, 3728–3731 (1993).
- Bodo, P. & Sundgren, J.-E. Titanium deposition onto ion-bombarded and plasma-treated polydimethylsiloxane: surface modification, interface, and adhesion. *Thin Solid Films* **136**, 147–159 (1986).
- Martin, G. C. *et al.* The metallization of silicone polymers in the rubbery and the glassy state. *J. Appl. Phys.* **53**, 797–799 (1982).
- Allen, H. G. *Analysis and Design of Structural Sandwich Panels* (Pergamon, New York, 1969).
- Hutchinson, J. W., Thouless, M. D. & Liniger, E. G. Growth and configuration stability of circular, buckling-driven film delaminations. *Acta Metall. Mater.* **40**, 295–308 (1992).
- Kumar, A., Abbott, N. L., Kim, E., Biebuyck, H. A. & Whitesides, G. M. Patterned self-assembled monolayers and meso-scale phenomena. *Acc. Chem. Res.* **28**, 219–226 (1995).
- Xia, Y. & Whitesides, G. M. Soft Lithography. *Angew. Chem. Int. Edn* **37**, 550–575 (1998).
- Ohring, M. *The Material Science of Thin Films* (Academic, San Diego, 1992).

Acknowledgements. We thank D. Munn for helpful discussions. This work was supported by the NSF, the Office of Naval Research, and the Defense Advanced Research Projects Agency. N.B. thanks the Department of Defense for a predoctoral fellowship.

Correspondence and requests for materials should be addressed to G.M.W. (e-mail: gwhitesides@gmgroup.harvard.edu).

Polymer gels with engineered environmentally responsive surface patterns

Zhibing Hu*, Yuanze Chen*, Changjie Wang*,
Yindong Zheng* & Yong Li†

* Department of Physics, University of North Texas, Denton, Texas 76203, USA

† Kimberly-Clark Corporation, Neenah, Wisconsin 54956, USA

The polymer gels called hydrogels may be induced to swell or shrink (taking up or expelling water between the crosslinked polymer chains) in response to a variety of environmental stimuli, such as changes in pH or temperature, or the presence of a specific chemical substrate¹. These gels are being explored for several technological applications, particularly as biomedical materials². When hydrogels swell or shrink, complex patterns may be generated on their surfaces^{3–7}. Here we report the synthesis and controlled modulation of engineered surface patterns on environmentally responsive hydrogels. We modify the character of a gel surface by selectively depositing another material using a mask. For example, we use sputter deposition to imprint the surface of an *N*-isopropylacrylamide (NIPA) gel with a square array of gold thin films. The periodicity of the array can be continuously varied as a function of temperature or electric field (which alter the gel's volume), and so such an array might serve as an optical grating for sensor applications. We also deposit small areas of an NIPA gel on the surface of an acrylamide gel; the patterned area can be rendered invisible reversibly by switching the temperature above or below the lower critical solution temperature of the NIPA gel. We anticipate that these surface patterning techniques may find applications in display and sensor technology.

We first discuss the surface patterns made by use of the sputtering-deposition technique. *N*-isopropylacrylamide (NIPA) gel slabs were made by free-radical polymerization, as follows. A mixture of 7.8 g of *N*-isopropylacrylamide, 133 mg of methylene-bis-acrylamide as a crosslinker, and tetramethylethylenediamine (240 μl) as an accelerator, was dissolved in 100 ml of deionized and distilled water. Nitrogen gas was bubbled through the solution to remove dissolved oxygen. The polymerization was initiated by adding 40 mg of ammonium persulphate. The samples were kept in water for several days to wash out chemical residues. The NIPA gels were then

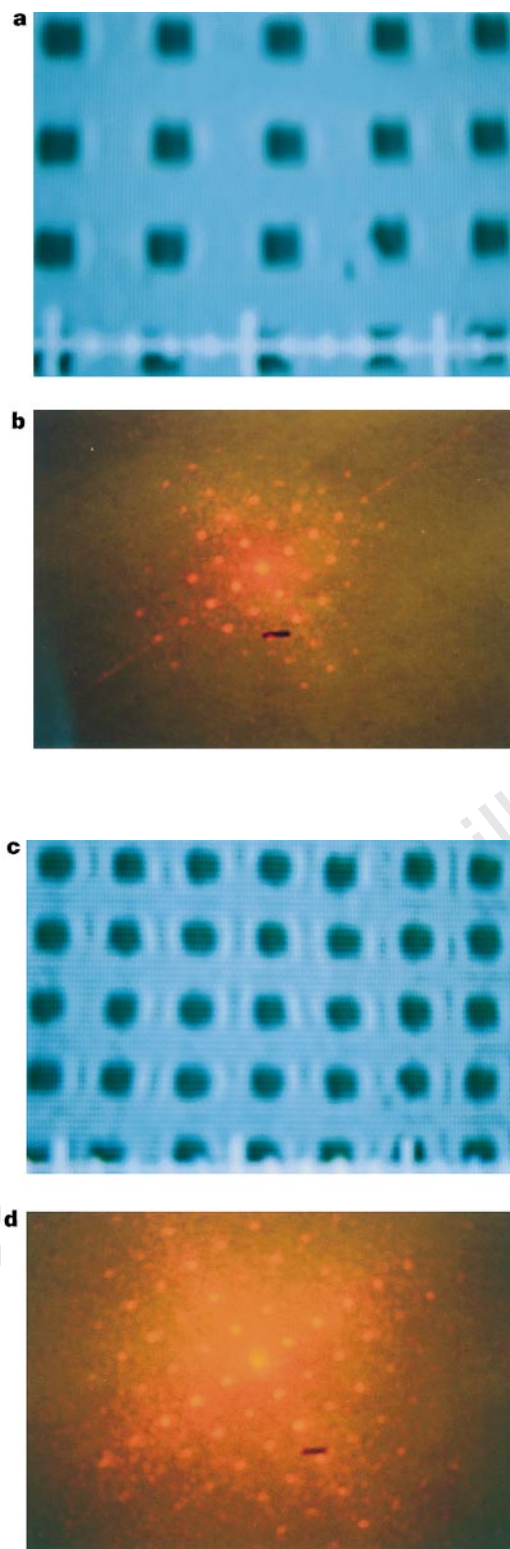


Figure 1 The periodic square surface array on the NIPA hydrogel and its diffraction pattern, at various temperatures. At 30°C: **a**, the surface array; **b**, the diffraction pattern. At 33.6°C: **c**, the surface array; **d**, the diffraction pattern. In the diffraction experiment, the distance between the screen and the sample was 95 cm. In **a** and **c**, each small division in the white scale at the bottom of the image is 0.0193 mm; in **b** and **d**, the black bar is 1 cm long.

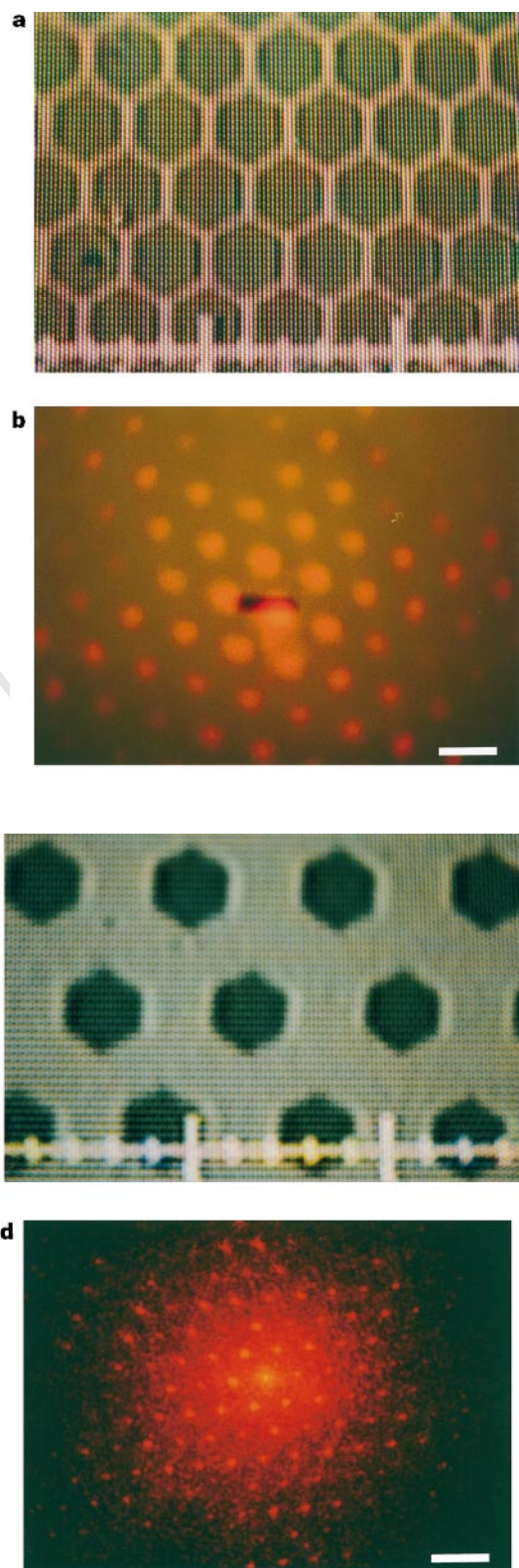


Figure 2 The periodic hexagonal surface array on the NIPA gel in the dehydrated and in the hydrated states at room temperature. **a**, The surface array of a dry NIPA gel film. **b**, The light diffraction pattern produced by the sample in **a**. **c**, The increased periodicity of the surface structure after immersion of the dry gel in water. **d**, The corresponding diffraction pattern from the sample of **c**. In the diffraction experiment, the distance between the screen and the sample was 100 cm. In **a** and **c**, each small division in the white scale at the bottom of the image is 0.0193 mm; in **b** and **d** the white bar is 2 cm long.

dried slowly in a partially sealed glass container so that they shrank uniformly.

SPI Slim Bar Grids (SPI supplies, West Chester, Pennsylvania) were used as masks for the sputter deposition. For the square mesh grid, the bar width was 6 μm and the hole width was 19 μm . A dry NIPA gel was covered by a grid and placed on the anode in the chamber of a sputter coater. During the coating operation, the chamber was pumped to a low-level vacuum with an argon gas pressure of 0.04 mbar. A sputtering voltage of 2.5 kV was applied to ionize the argon. When the argon ions struck the gold cathode, gold atoms were ejected and impinged upon the sample through the openings of the mask. This process was continued for ~ 3.5 min, giving a gold thickness of ~ 53 nm. After coating, the mask was removed and the gel was immersed in water. The affinity between gold and the polymer gel may be due to the interaction between the lone-pair electron of the amine group in the NIPA gel and the empty orbitals of gold⁸, and to entrapment of gold in the gel. The temperature of samples was controlled by a circulation water bath. The sample surfaces were examined by an optical microscope, and the diffraction experiment was performed by placing the gel sample between a He–Ne laser and a screen. The microscopic structure and the diffraction patterns were recorded using a camera.

Figure 1a shows the surface array of a swollen NIPA gel film in water at 30 °C. The dark area is the gold-covered surface and the light area is the bare gel surface. The spaces between the small gold squares in the gel surface can serve as slits for light diffraction. Because such slits are aligned in both the x - and y -directions, the

diffraction pattern should consist of bright and dark spots instead of fringes when monochromatic light passes through these slits. Such a diffraction pattern is shown in Fig. 1b.

As the temperature was increased from room temperature, the NIPA gel shrank owing to the volume phase transition⁴. This caused the array periodic constant (d) to decrease. The periodic surface array and its diffraction pattern for the gel at 33.6 °C are shown in Fig. 1c and d, respectively. The gel was thermally cycled at least seven times, and the temperature-induced change in the diffraction pattern was observed to be reversible. Several different samples were synthesized and tested, and their diffraction patterns were also reproducible. We have also successfully ‘tuned’ the periodicity of the surface array of an ionic NIPA gel by applying an electric field. The electric field caused migration of positive counter ions and water inside the gel towards the negative electrode, and resulted in volume shrinkage⁹.

Although the surface arrays discussed here have a periodic square structure, many other structural arrays may be synthesized by using different masks; these include strip, rectangular, hexagonal, and even non-periodic surface arrays. Figure 2a shows a hexagonal surface array on a dehydrated NIPA gel at room temperature. The corresponding diffraction pattern is shown in Fig. 2b and exhibits a clear six-fold symmetry. When the gel is fully swollen in water at room temperature, the periodicity of the surface array is significantly larger, as is clearly demonstrated by both optical (Fig. 2c) and diffraction (Fig. 2d) observations.

The periodic array produced on the gel surface can serve as a two-dimensional grating, and may be compared with a previous study of three-dimensional diffraction from a crystalline colloidal array of polystyrene spheres embedded within a gel¹⁰. The width of the slits and the number of the slits per unit area of the gel surface grating can be changed by external stimuli such as temperature and electric field. The location of the diffraction bright spots is described by $\sin\theta_x = n\lambda/d_x$ and $\sin\theta_y = m\lambda/d_y$; here subscripts x and y represent the horizontal and the vertical direction, respectively, θ is the angle relative to the central spot, λ is the wavelength (633 nm for the He–Ne laser), n and m are the order of diffraction, and d is the grating constant or periodicity. As the periodicity of the gel surface pattern changes in response to the external stimuli, the diffraction angle should change accordingly. Such a gel grating has potential for applications in optical filters and sensors, and in optical communications.

The gel with periodic surface patterns can be used as a sensor. For example, the deformation of a gel under external force can be easily monitored using such a gel. In one experiment, we suspended in water a gel with a periodic surface array, and subjected it to a uniaxial force at its bottom end along the y -direction. Diffraction patterns of the gel were recorded for various loads. For zero force, the separation (Δl_x) between neighbouring diffraction spots along the x -direction was equal to that (Δl_y) along the y -direction. The ratio of $\Delta l_y/\Delta l_x$ decreased, however, as the external force increased. For the load of 5.85 g, $\Delta l_y/\Delta l_x \approx 0.72$.

This incorporation of a metal array on a hydrogel surface opens a new avenue to the manufacture of microelectrode array devices which can greatly enhance sensor performance¹¹. The combination of microelectrode arrays and ‘smart’ gels has potential applications in ion chromatography¹¹, the detection of enzyme activity¹², and the monitoring of cell electrical activity¹³.

We now consider the synthesis of an engineered surface pattern; this was accomplished by depositing one polymer gel (NIPA) on the surface of another polymer gel (polyacrylamide, PAAM, gel). The PAAM gel was made using the NIPA recipe but with 7.8 g of *N*-isopropylacrylamide replaced by 5 g of acrylamide, the use of 13.3 mg rather than 40 mg of ammonium persulphate, and the addition of 4 mg riboflavin-5'-phosphate (ultraviolet photoinitiator). The NIPA pre-gel solution was put on the surface of the PAAM gel; the sample was then covered by a mask (we used an outline map

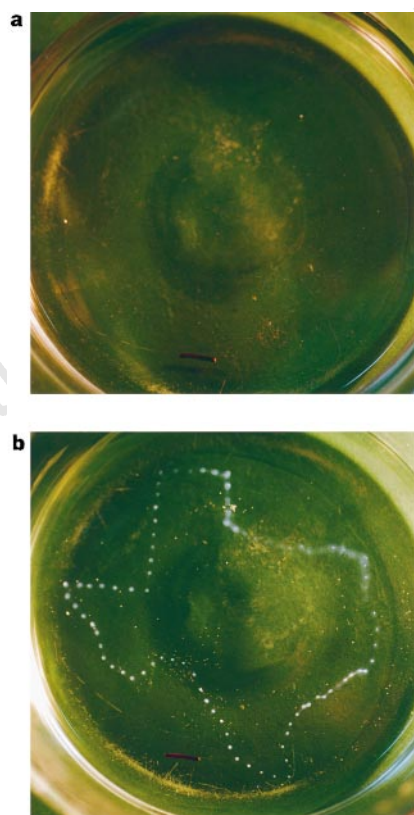


Figure 3 Temperature-responsive surface pattern in the NIPA-deposited PAAM gel. **a**, At room temperature, the NIPA-deposited PAAM gel is transparent and no surface pattern is seen. **b**, When heated to above 37 °C, the outline image of Texas appears. This is because at >37 °C the polymerized NIPA-deposited areas become cloudy while the PAAM gel remains transparent. The whole process takes ~ 10 s, and after the sample is cooled down, the map disappears: the process is fully reversible. Scale bar, 1 cm.

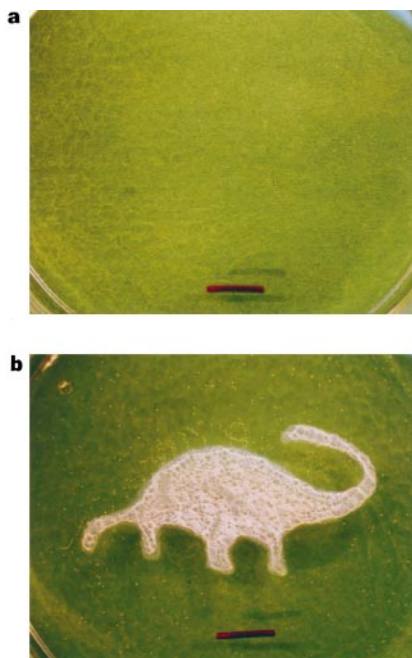


Figure 4 A gel display with a solid image. **a**, At room temperature, the NIPA-deposited PAAM gel is transparent and no surface pattern is seen. **b**, When heated to $>37^{\circ}\text{C}$, the solid image of a dinosaur appears. Scale bar, 1 cm.

of Texas) and illuminated by ultraviolet light for ~ 60 min. The NIPA solution gelled only at areas that were exposed to ultraviolet light; at these areas, an interpenetrating polymer network¹⁴ of NIPA and PAAM was formed, with an estimated depth of ~ 0.5 mm.

As both the PAAM substrate and the NIPA gel on the surface of the substrate are transparent at room temperature, no surface structure can be observed (Fig. 3a). When this surface-engineered sample was warmed up to 37°C , the areas of polymerized NIPA become cloudy^{15,16} while the PAAM gel remained in the transparent state. As a result, the outline image of a Texas map appears at the surface of the PAAM gel (Fig. 3b). This map can be turned on or turned off by simply 'switching' the temperature above or below the lower critical solution temperature, T_c ($= 34^{\circ}\text{C}$) of the NIPA gel. The switching time was ~ 10 s, much shorter than that required for a shape memory gel¹⁴.

To further demonstrate the potential of the gels with engineered surface patterns, a gel display consisting of a PAAM substrate with the NIPA gel filling the unmasked surface area was made, and is shown immersed in water in Fig. 4. At room temperature (Fig. 4a), both NIPA and PAAM gels were transparent and no pattern was observed. At 37°C (Fig. 4b), the NIPA-deposited area becomes cloudy, and results in appearance of a solid image; a representation of a dinosaur. The gel display developed here is responsive to temperature, but by changing the chemical composition of the gels, the gel displays could be made responsive to other external stimuli such as salt concentration and infrared light. The image of the gel display can be clearly seen at any angle, in contrast to a well established liquid-crystal display that requires head-on viewing angles. □

Received 20 November 1997; accepted 24 February 1998.

1. Tanaka, T. Gels. *Sci. Am.* **244**, 124–138 (1981).
2. Peppas, N. A. & Langer, R. New challenges in biomaterials. *Science* **263**, 1715–1720 (1994).
3. Tanaka, T. *et al.* Mechanical instability of gels at the phase transition. *Nature* **325**, 796–798 (1987).
4. Sato-Matsuo, E. & Tanaka, T. Patterns in shrinking gels. *Nature* **358**, 482–485 (1992).
5. Li, Y., Li, C. & Hu, Z. Pattern formation of constrained acrylamide/sodium acrylate copolymer gels in acetone/water mixture. *J. Chem. Phys.* **100**, 4637–4644 (1994).
6. Li, C., Hu, Z. & Li, Y. Temperature and time dependence of patterns of surface patterns in constrained ionic N-isopropylacrylamide gels. *J. Chem. Phys.* **100**, 4645–4652 (1994).

7. Suzuki, A., Yamazaki, M. & Kobiki, Y. Direct observation of polymer gel surfaces by atomic force microscopy. *J. Chem. Phys.* **104**, 1751–1757 (1996).
8. Grabar, K. C., Freeman, G. G., Hommer, M. B. & Natan, M. J. Preparation and characterization of Au colloid monolayers. *Anal. Chem.* **67**, 735–743 (1995).
9. Hirotsu, S. Electric-field-induced phase transition in polymer gels. *Jpn J. Appl. Phys.* **24**, 396–398 (1985).
10. Weissman, J. M., Sunkara, H. B., Tse, A. S. & Asher, S. A. Thermally switchable periodicities and diffraction from mesoscopically ordered materials. *Science* **274**, 959–960 (1996).
11. Imisides, M. D., John, R. & Wallace, G. G. Microsensors based on conducting polymers. *Chemtech* **26**, 19–25 (1996).
12. Wang, J. & Chen, Q. Enzyme microelectrode array strips for glucose and lactate. *Anal. Chem.* **66**, 1007–1011 (1994).
13. Misch, W., Bock, J., Egert, U., Hammerle, H. & Mohr, A. A thin film microelectrode array for monitoring extracellular neuronal activity in vitro. *Biosensors Bioelectr.* **9**, 737–741 (1994).
14. Hu, Z., Zhang, X. & Li, Y. Synthesis and application of modulated polymer gels. *Science* **269**, 525–527 (1995).
15. Li, Y., Wang, G. & Hu, Z. Turbidity study of spinodal decomposition of N-isopropylacrylamide gels. *Macromolecules* **28**, 4194–4197 (1995).
16. Bansil, R., Liao, G. & Falus, P. Kinetics of spinodal decomposition in chemically crosslinked gels. *Physica A* **231**, 346–358 (1996).

Acknowledgements. We thank the Donors of the Petroleum Research Fund, administered by the American Chemical Society, and the US Army Research Office for support of this work.

Host-guest encapsulation of materials by assembled virus protein cages

Trevor Douglas* & Mark Young†

* Department of Chemistry, Temple University, Philadelphia, Pennsylvania 19122-2585, USA

† Department of Plant Pathology, Montana State University, Bozeman, Montana 59717-0314, USA

Self-assembled cage structures of nanometre dimensions can be used as constrained environments for the preparation of nanostructured materials^{1,2} and the encapsulation of guest molecules³, with potential applications in drug delivery⁴ and catalysis⁵. In synthetic systems the number of subunits contributing to cage structures is typically rather small^{3,6}. But the protein coats of viruses (virions) commonly comprise hundreds of subunits that self-assemble into a cage for transporting viral nucleic acids. Many virions, moreover, can undergo reversible structural changes that open or close gated pores to allow switchable access to their interior⁷. Here we show that such a virion—that of the cowpea chlorotic mottle virus—can be used as a host for the synthesis of materials. We report the mineralization of two polyoxometalate species (paratungstate and decavanadate) and the encapsulation of an anionic polymer inside this virion, controlled by pH-dependent gating of the virion's pores. The diversity in size and shape of such virus particles make this a versatile strategy for materials synthesis and molecular entrapment.

We have used the well defined cowpea chlorotic mottle virus (CCMV)⁷ as a model system for reversibly gated entrapment of inorganic minerals and an organic polymer. CCMV virions are 280 Å in diameter and the protein shell defines an inner cavity ~ 180 Å in diameter as measured by electron microscopy. This virus is composed of 180 identical coat protein subunits arranged on an icosahedral lattice. Purified viral coat protein subunits can be easily assembled *in vitro* into empty virion particles⁸. The dimensions of the virion cavity define the upper limit for crystal growth of the entrapped mineral. In addition, each coat protein subunit presents at least nine basic residues (arginine and lysine) to the interior of the cavity⁹. This creates a positively charged interior cavity surface, which provides an interface for inorganic crystal nucleation and growth. From the structure of the CCMV particle it is known that the outer surface of the protein shell is not highly charged⁷, thus the inner and outer surfaces of this molecular container provide chemically unique environments which we have exploited to spatially localize and control the mineralization reaction.

In their native state, viruses are protein assemblies which act as host containers for nucleic acid storage and transport. We have subverted this natural function and used the assembled cage-like structure of the virion for the formation and entrapment of both inorganic and organic polymer species; materials which are useful in such wide-ranging areas as catalysis¹⁰ and drugs^{11,12}.

Previous studies have shown that supramolecular protein assemblies, derived from the iron storage protein ferritin, can be used to spatially constrain mineralization reactions^{13,14}. However, the ferritin protein assembly is limited in scope to a single size and shape¹⁵. On the other hand, viruses occur in a wide range of size and morphology and many examples of polymorphic viruses exist^{16,17}. For example, poliovirus exists as a 30-nm-diameter sphere¹⁸ while the tobacco mosaic virus forms stable 300-nm rod-like structures¹⁹ and CCMV is known to exist in many architecturally distinct polymorphs^{16,20}. These complex cage-like assemblies, devoid of their nucleic acid, provide cavities of well defined size, shape and chemical environment and can all be viewed as potential hosts for size- and shape-constrained molecular entrapment and materials synthesis.

Dynamic structural transitions occur in many virions¹⁶, induced by defined chemical switches, which provide unique molecular gating mechanisms to control the containment and release of entrapped materials. CCMV undergoes a reversible pH-dependent swelling which results in a 10% increase in virus dimension⁷. Structural analysis has revealed that this transition is the result of an expansion at the pseudo three-fold axis of the virus particle which causes the formation of 60 separate openings (each ~20 Å in diameter) in the protein shell (Fig. 1). Under swollen conditions (pH > 6.5) these openings allow free molecular exchange between the virus cavity and the bulk medium. In contrast, in the non-swollen form (pH < 6.5) there is no apparent exchange of large molecules between the cavity and the bulk medium. This structural transition allows the virus particle to selectively entrap and release materials from within the central cavity by a mechanism analogous to reversible gating described in synthetic host-guest systems²¹.

We have coupled this pH-dependent gating mechanism of the host virion to a pH-dependent inorganic oligomerization reaction in order to load, crystallize, and entrap mineral particles of well defined size (Fig. 2). The oligomerization of aqueous molecular tungstate (WO_4^{2-}) species to form paratungstate ($\text{H}_2\text{W}_{12}\text{O}_{42}^{10-}$) polyanions is induced by lowering the pH (ref. 22). The paratungstate polyanions have significantly lower solubility than the precursor tungstate ions. The empty virions were incubated with the inorganic precursor ions under conditions (pH > 6.5) where the virus exists in its open (swollen) form and allows all ions access to the central cavity. After incubation, the pH was lowered to a point where the virion undergoes a structural transition and the pores in the protein

shell close (Fig. 1). The virion gating, coupled to the tungstate oligomerization, results in the spatially selective crystallization and entrapment of the paratungstate ion within the virion. Crystallization occurred only within the virus particle and no bulk mineralization was observed in virion-free controls or in solutions containing only assembled virions but without tungstate.

The mineralized virions were purified by centrifugation on sucrose gradients giving an indication of the stability of guest (mineral) incarceration complex. Purified samples were subsequently examined by transmission electron microscopy (TEM) and the electron-dense mineral cores were observed to be commensurate in size and shape with the internal diameter of the virus particle (average 150 Å, Fig. 3a). This process is clearly a crystallization, not just an entrapment, because the size of the crystals formed are larger than expected (67 Å diameter) for entrapped material only based on crystal density²³. Negative stain of these samples showed the presence of the intact capsid protein shell surrounding these cores (Fig. 3b). This result indicated the spatial selectivity of the mineralization. Selected-area diffraction from a

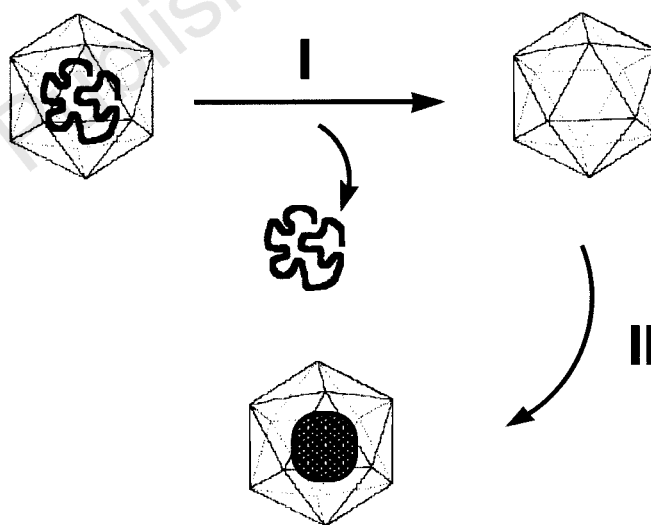


Figure 2 Schematic illustration of the synthetic approach for mineralization within the virus particle. Step I involves the removal of viral RNA and purification of the empty virus particle by ultracentrifugation of sucrose gradients. Step II involves the selective mineralization of an inorganic paratungstate species within the confines of the virus particle, at pH 6.5 and 6°C. Transmission electron microscopy of the virion mineralized with decavanadate showed nearly identical cores.

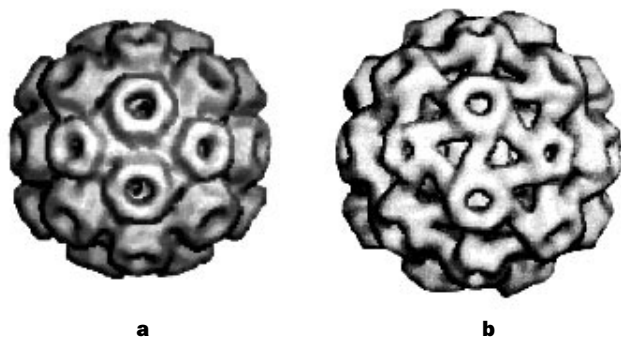


Figure 1 Cryo electron microscopy and image reconstruction of the cowpea chlorotic mottle virus (CCMV). **a**, in an unswollen condition induced by low pH; **b**, in a swollen condition induced by high pH (images provided by T. Baker, Purdue Univ.). Swelling at the pseudo-three-fold axis results in the formation of 60 20-Å pores.

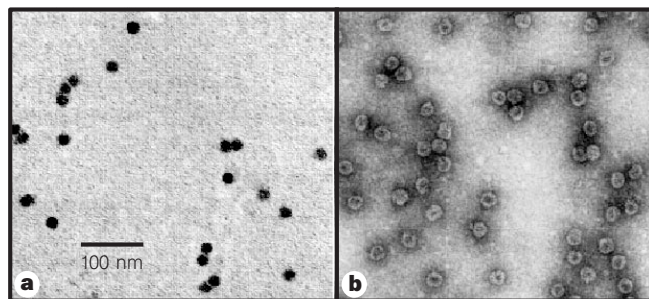


Figure 3 TEM images of paratungstate-mineralized virus particles after isolation by centrifugation on a sucrose gradient. **a**, An unstained sample showing discrete electron dense cores; **b**, a negatively stained sample of **a** showing the mineral core surrounded by the intact virus protein cage. Scale bar, 100 nm.

single core yielded a diffraction pattern consistent with the ammonium salt of the paradodecatungstate ion $(\text{NH}_4)_{10}\text{H}_2\text{W}_{12}\text{O}_{42}\cdot 4\text{H}_2\text{O}$ (ref. 23) and high-resolution TEM of individual cores, still enclosed within the virion, revealed a uniform material with continuous 2.8-Å lattice fringes (050) illustrating the single-crystal nature of the mineral (Fig. 4b). Control reactions in the absence of virus revealed no electron-dense cores, as did virus incubated in the absence of any tungstate. Bulk crystallization induced by slow evaporation of solvent, at 6 °C, yielded plate-like crystals of $\text{Na}_2(\text{NH}_4)_8\text{H}_2\text{W}_{12}\text{O}_{42}\cdot 12\text{H}_2\text{O}$ (ref. 24) with chemical analysis of Na (1.26%), W (67.34%), NH_4^+ (3.34%); by X-ray powder diffraction the following planes were strongly observed (020) 11.3 Å, (040) 5.8 Å, (060) 3.85 Å, (080) 2.9 Å.

As a further example of the general use of this method, a similar approach was used to constrain the mineralization of a vanadate species ($\text{V}_{10}\text{O}_{28}^{6-}$) within the virion of CCMV. This synthesis also results in the formation of monodisperse nanoparticles with essentially the same size distribution as seen for the paratungstate (Fig. 4a and b). Control reactions in the absence of the virion yielded yellow-orange crystals after slow evaporation of water at 6 °C. These crystals showed X-ray powder diffraction consistent with a mixture of $\text{Na}_6\text{V}_{10}\text{O}_{28}\cdot 18\text{H}_2\text{O}$ (*d* spacings (in Å) 10.64, 10.21, 8.0, 7.46, 5.86, 3.45, 2.87 (ref. 25) and $(\text{NH}_4)_6\text{V}_{10}\text{O}_{28}\cdot 18\text{H}_2\text{O}$ (*d* spacings 8.41, 8.0, 6.25, 6.19, 3.16 (ref. 26).

Crystallization of the polyoxometalate minerals occurs in a spatially selective manner within the virus cavity. The reaction conditions used here have been tailored so that the host gating and guest oligomerization complement each other in order to entrap the resultant mineral. We believe that the polyoxometalate crystallization is electrostatically induced at the interior surface of the protein surrounding the cavity. From the structure of CCMV it is known that at least 1,620 basic residues are localized on this interior surface⁷. Thus, the negatively charged paratungstate and decavanadate ions ($\text{H}_2\text{W}_{12}\text{O}_{42}^{10-}$, $\text{V}_{10}\text{O}_{28}^{6-}$) might reasonably be expected to aggregate at this interface, and facilitate the crystal nucleation event²⁷. Therefore, we suggest that the protein shell acts as a spatially selective nucleation catalyst in the paratungstate

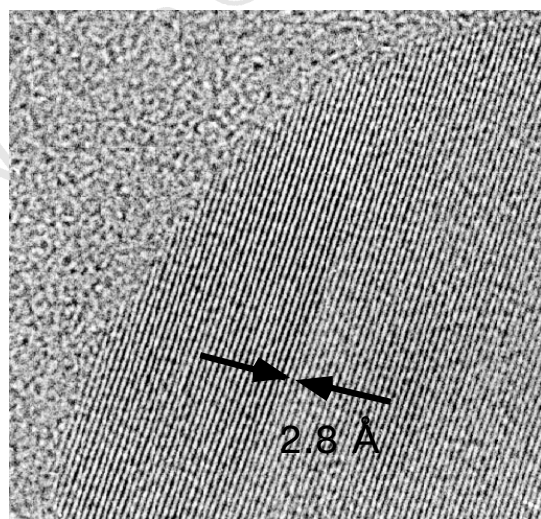


Figure 4 High-resolution TEM image of a corner of an individual paratungstate core, encapsulated within the virion. The single-crystal nature of the material is illustrated. The lattice fringe (050) was measured at 2.8 Å.

mineralization, in addition to its role as a size- and shape-constrained reaction vessel. This is consistent with the notion that carcerand interiors provide a “new phase of matter”²⁸ and are capable of catalysing chemical reactions⁵. This activity is also consistent with the crystal nucleation process which occurs in the iron storage protein ferritin²⁹.

To investigate the electrostatic aspects of this host–guest interaction and the influence of gating on the entrapment, we undertook the pH-dependent encapsulation of an anionic organic polymer (poly-anetholesulphonic acid). The polymer was incubated with the empty virion at high pH (7.5) followed by a lowering of the pH to well below the gating threshold (pH 4.5). This resulted in selective encapsulation of the polymer. The polymer-loaded virion was isolated by gradient centrifugation (Fig. 5a) and imaged by TEM, revealing the intact protein shell surrounding an electron-transparent core (that is, no stain intrusion into the virion interior) (Fig. 5b). Stain intrusion is prevented by the presence of the organic polymer. In addition, the isolated material showed the characteristic ultraviolet-visible spectrum of the aromatic anionic polymer. Control reactions in which the empty virions were incubated at low pH (under non-swollen conditions) yielded no measurable uptake of the polymer.

This approach to encapsulation chemistry has clear applications to the formation of other nanophase materials as well as for molecular encapsulation. Whereas the protein cages of virions have previously been used as delivery systems for non-native nucleic acids³⁰, the work presented here represents an entirely synthetic manipulation of the cage-like properties of the virion. The range of viral morphology and size allow great flexibility for adapting this methodology to control the size and shape of the entrapped material, which is limited only by access to the protein interior

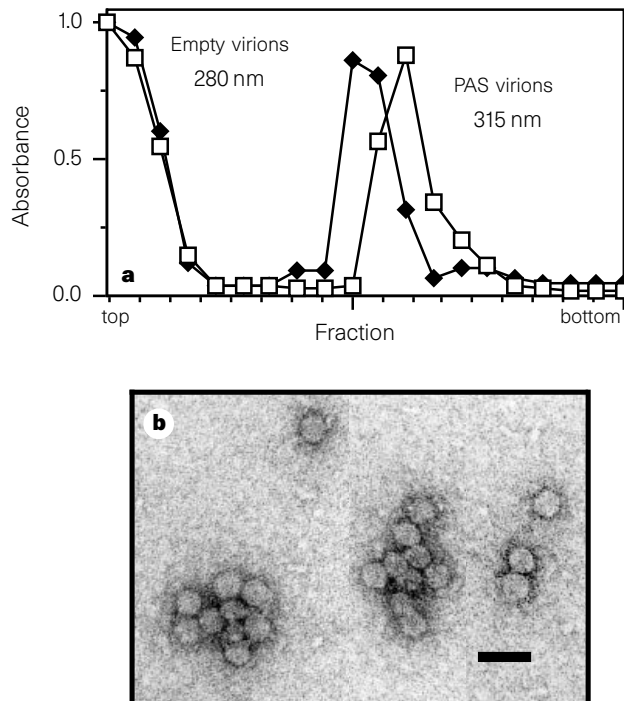


Figure 5 Encapsulation of polyanetholesulphonic acid (PAS) within the CCMV virion. **a**, Absorption profile of the relative sedimentation of empty virions (comprising only protein monitored at 280 nm) and of PAS-containing virions (monitored at both 280 nm and a unique 315-nm absorbance), on 10–40% sucrose gradients. Absorbances have been normalized. **b**, Composite TEM image of negatively stained PAS-containing virions, showing no stain intrusion into the interior of the virion due to the presence of the encapsulated PAS. Scale bar, 50 nm.

and could include inorganic and organic species. In addition, the protein cage could be easily and routinely modified by design. The electrostatic environment within the cavity could be altered by site-directed mutagenesis to induce additional specific interactions, and the outer surface of the protein could be modified to facilitate specific biological targeting and surface interaction. The demonstration of gating illustrates a mechanism for molecular capture and release not previously exploited. □

Methods

Self assembly of virions and purification. Empty virions were assembled from purified coat protein as previously described^{8,31}. The empty virus particles were assembled by dialysis of a 1.0 mg ml⁻¹ coat protein suspension against buffer (0.1 M NaOAc, pH 4.8, 0.9 M NaCl, 0.2 mM PMSF) for 10–14 h at 4 °C. The assembly mixtures were loaded on to 10–40% sucrose gradients and centrifuged at 38,000 r.p.m., 4 °C for 3.0 h in a Beckman SW41ti rotor. The gradients were fractionated on an ISCO model 640 gradient fractionator while monitoring absorption at 280 nm. The desired fractions were collected and dialysed against buffer (0.05 M Tris HCl pH 7.0).

Mineralization of the virion

Paratungstate. The empty virions (50 µg) were incubated with inorganic precursor ions (75 mM Na₂WO₄, 50 mM NH₄Cl, pH 6.5) for 48 h at 6 °C. Under these conditions the virus exists in its open (swollen) form and allows all ions access to the central cavity. After incubation the virus solution was concentrated using ultracentrifuge membrane with a relative molecular mass cut-off of 100K, and extensively washed with acetate buffer (0.1 M) at pH 5.0. The mineralized virion was purified by ultracentrifugation on a 10–40% sucrose gradient. Bulk crystallization (in the absence of virion) was induced by slow evaporation of solvent at 6 °C and yielded plate-like crystals. X-ray powder diffraction was measured on a Scintag X1 advanced diffraction system, and chemical analysis was performed at Energy Laboratories (Billings, MT).

Decavanadate. The empty virions (50 µg) were incubated with 160 mM NaVO₃ at pH 6.5 for 1 h. An equal volume of 100 mM NH₄Cl, pH 3.0 was added resulting in the formation of a bright yellow-orange solution containing the V₁₀O₂₈⁶⁻ anion. After 24 h at 6 °C the reaction mixture was concentrated and washed extensively with NH₄Cl (pH 4.0), using ultracentrifugation. Bulk crystallization (in the absence of virion) was induced by slow evaporation of solvent at 6 °C and yielded deep red-orange crystals. X-ray powder diffraction was measured on a Rigaku D/max diffractometer.

Polyanetholesulphonic acid. Empty virions (100 µg) were incubated in a 20 mg ml⁻¹ polyanetholesulphonic acid (Sigma Chemical, average relative molecular mass 9K–11K), 50 mM Tris-HCl pH 7.5 solution at 25 °C. After incubation, the pH of the solution was decreased to pH 4.5 with 100 mM sodium acetate buffer and washed extensively with acetate buffer in a Centricon 100. The sample was subsequently loaded on to, and centrifuged through, a 10–40% sucrose gradient as described above. Gradient fractions were collected and monitored by absorption at 280 nm and 315 nm. Gradient fractions were washed and concentrated using Centricon 100s. Samples were examined by TEM. The strong, unique absorption characteristic of polyanetholesulphonic acid (315 nm) was only present in virion-containing fractions. The controls (initially incubating an equivalent amount of empty virions at pH 4.5 instead of pH 7.5) did not show absorption characteristic of polyanetholesulphonic acid.

Transmission electron microscopy. Liquid samples were placed on TEM grids (carbon-coated, formvar-covered, Cu) and excess solution removed with filter paper. Samples were imaged at 80 kV using a Zeiss EM 10CA or a Philips 400T and high-resolution electron microscopy was performed on a JEOL 4000EX operating at 400 kV. Virions were negatively stained using uranyl acetate.

Received 22 October 1997; accepted 11 February 1998.

- Mann, S. *Biomimetic Materials Chemistry* (VCH, New York, 1996).
- Stucky, G. D. & MacDougall, J. E. Quantum confinement and host/guest chemistry: Probing a new dimension. *Science* **247**, 669–678 (1990).
- Rebek, J. J. Assembly and encapsulation with self complementary molecules. *Chem. Soc. Rev.* **25**, 255–264 (1996).
- Park, K. *Controlled Drug Delivery: Challenges and Strategies* (Am. Chem. Soc., Washington DC, 1997).
- Kang, J. & Rebek, J. J. Acceleration of a Diels-Alder reaction by a self-assembled molecular capsule. *Nature* **385**, 50–52 (1997).
- MacGillivray, L. R. & Atwood, J. L. A chiral spherical molecular assembly held together by 60 hydrogen bonds. *Nature* **389**, 469–472 (1997).

- Speir, J. A., Munshi, S., Wang, G., Baker, T. S. & Johnson, J. E. Structures of the native and swollen forms of cowpea chlorotic mottle virus determined by X-ray crystallography and cryo-electron microscopy. *Structure* **3**, 63–78 (1995).
- Zhao, X., Fox, J. M., Olson, N. H., Baker, T. S. & Young, M. J. *In vitro* assembly of cowpea chlorotic mottle virus from coat protein expressed in *Escherichia coli* and *in vitro*-transcribed viral cDNA. *Virology* **207**, 486–494 (1995).
- van der Graaf, M., van Mierlo, C. P. M. & Hemminga, M. A. Solution conformation of a peptide fragment representing a proposed RNA-binding site of a viral coat protein studied by two-dimensional NMR. *Biochemistry* **30**, 5722–5727 (1991).
- Misono, M. in *Polyoxometalates: From Platonic Solids to Anti-Retroviral Activity* (eds Pope, M. T. & Müller, A.) 255–266 (Kluwer Academic, Dordrecht, 1994).
- Sarafianos, S. G., Kortz, U., Pope, M. T. & Modak, M. J. Mechanism of polyoxometalate-mediated inactivation of DNA polymerases: an analysis with HIV-1 reverse transcriptase indicates specificity for the DNA-binding cleft. *Biochem. J.* **319**, 619–626 (1996).
- Belford, D. A., Hendry, I. A. & Parish, C. R. Investigation of the ability of several naturally occurring and synthetic polyanions to bind to and potentiate the biological activity of acidic fibroblast growth factor. *J. Cell Physiol.* **157**, 184–189 (1993).
- Meldrum, F. C., Wade, V. J., Nimmo, D. L., Heywood, B. R. & Mann, S. Synthesis of inorganic nanophase materials in supramolecular protein cages. *Nature* **349**, 684–687 (1991).
- Meldrum, F. C., Heywood, B. R. & Mann, S. Magnetoferritin: In vitro synthesis of a novel magnetic protein. *Science* **257**, 522–523 (1992).
- Douglas, T. in *Biomimetic Approaches in Materials Science* (ed. Mann, S.) 91–115 (VCH, New York, 1996).
- Johnson, J. E. & Speir, J. Quasi-equivalent viruses: A paradigm for protein assemblies. *J. Mol. Biol.* **269**, 665–675 (1997).
- Chiu, W., Burnett, R. & Garcea, R. *Structural Biology of Viruses* (Oxford Univ. Press, 1997).
- Hogle, J. M., Chow, M. & Filman, D. J. Three-dimensional structure of poliovirus at 2.9 Å resolution. *Science* **229**, 1358–1365 (1985).
- Namba, K., Pattanayek, R. & Stubbs, G. Visualization of protein-nucleic acid interactions in a virus; Refinement of intact tobacco mosaic virus structure at 2.9 Å resolution by fiber diffraction. *J. Mol. Biol.* **208**, 307–325 (1989).
- Bancroft, J. B., Wagner, G. W. & Bracket, C. E. The self-assembly of a nucleic-acid free pseudo-top component for a small spherical virus. *Virology* **36**, 146–149 (1968).
- Houk, K. N., Nakamura, K., Sheu, C. & Keating, A. E. Gating as a control element in constrictive binding and guest release by hemiacerands. *Science* **273**, 627–629 (1996).
- Pope, M. T. *Heteropoly and Isopoly Oxometalates* (Springer, Heidelberg, 1983).
- Evans, H. T. & Prince, E. Location of internal hydrogen atoms in the paratungstate polyanion by neutron diffraction. *J. Am. Chem. Soc.* **105**, 4838–4839 (1983).
- d'Amour, H. & Allman, R. Die kristallstruktur des dinatrium-oktammmonium-parawolframat-dodekahydrats, Na₂(NH₄)₈[H₂W₁₂O₄₂]·12H₂O. *Z. Kristallogr.* **138**, 5–18 (1972).
- Powder Diffraction File 20-1176 (Joint Committee on Powder Diffraction Standards, Swarthmore, PA).
- Johnson, G. K. & Murmann, R. K. Sodium and ammonium decavanadates (V). *Inorg. Synth.* **19**, 140–145 (1979).
- Heywood, B. R. & Mann, S. Template directed nucleation and growth of inorganic materials. *Adv. Mater.* **6**, 9–20 (1994).
- Sherman, J. C. & Cram, D. J. Carcer and interiors provide a new phase of matter. *J. Am. Chem. Soc.* **111**, 4527–4528 (1989).
- Lawson, D. M. et al. Solving the structure of human H ferritin by genetically engineering intermolecular crystal contacts. *Nature* **349**, 541–544 (1991).
- Smith, A. E. Viral vectors in gene therapy. *Annu. Rev. Microbiol.* **49**, 807–838 (1995).
- Fox, J. M., Zhao, X., Speir, J. A. & Young, M. J. Analysis of a salt stable mutant of cowpea chlorotic mottle virus. *Virology* **222**, 115–122 (1996).

Acknowledgements. We thank S. Brumfield, D. Mogk and G. Myer for experimental assistance, and R. Lakis for assistance with the electron microscopy.

Correspondence and requests for materials should be addressed to T.D. (e-mail: tdouglas@nimbus.ocis.temple.edu) or M.Y. (e-mail: uplmy@gemini.ocsc.montana.edu).

Possible triggering of Heinrich events by ice-load-induced earthquakes

A. G. Hunt & P. E. Malin

Division of Earth and Ocean Sciences, Nicholas School of the Environment, Duke University, Box 90235, Durham, North Carolina 27708-0235, USA

North Atlantic sediments dating from the last ice age contain layers of rock fragments from northeastern Canada (so-called Heinrich layers)¹. Like modern iceberg-borne sediments from Greenland, these layers have been attributed to ice-rafting episodes^{1–3}. Six Heinrich layers have been documented and correlated with climate changes^{4–8}. The layers, which are several centimetres thick, contain negligible amounts of foraminifera (which accumulate at a few millimetres per century), implying that they were deposited over just a few years. These ice-rafting Heinrich events are separated by progressively shorter intervals from about 40 to 6 kyr (ref. 9), and it has been suggested¹⁰ that they are related to the Milankovitch cycles in the Earth's orbital

parameters¹¹. Alternatively, they may be generated by forcing mechanisms arising from the internal dynamics of the Laurentide ice sheet¹². Here we suggest the possibility that the Heinrich events were precipitated by ice-load-induced earthquakes, analogous to those produced by reservoir water loads¹³. We suggest that near its edge, the Laurentide ice sheet sheared the Earth's crust, inducing repeated failure that released the ice rafts. This region (along Canada's northeastern seaboard) shows evidence of both current^{14,15} and past seismic activity owing to postglacial rebound. Our model accounts for the intervals between both the Heinrich events and the evidence of palaeoseismicity, and can be tested by studying local sedimentary relationships.

That changing ice loads can cause crustal failures not related to plate tectonics is supported by recently discovered postglacial earthquake faulting in Fennoscandia^{16,17}. As direct evidence for similar postglacial earthquake activity in northeastern Canada, we cite uplifted shorelines on Baffin Island and Newfoundland^{18,19}. These shorelines lie near offshore belts of seismicity that do not coincide with plate boundaries^{14,15} (Fig. 1). Instead, the seismicity lies along the edges of the former Laurentide ice sheet, in the areas where the Heinrich events originated (Fig. 2).

Using a simple model of the crust and mantle beneath the expanding and contracting ice sheet (Fig. 2), we show that the strains and strain rates around the sheet's edges match those of tectonic crustal failure, resulting in massive elastic rebound and ice-sheet damage there²⁰. By using these strains and strain rates, it seems possible to explain the time intervals of both the Heinrich events and the uplifted postglacial shorelines in terms of earthquake recurrence intervals (Fig. 3; Table 1). We note that our model can be tested, for example, by palaeoseismic and dating studies near the edges of the ice.

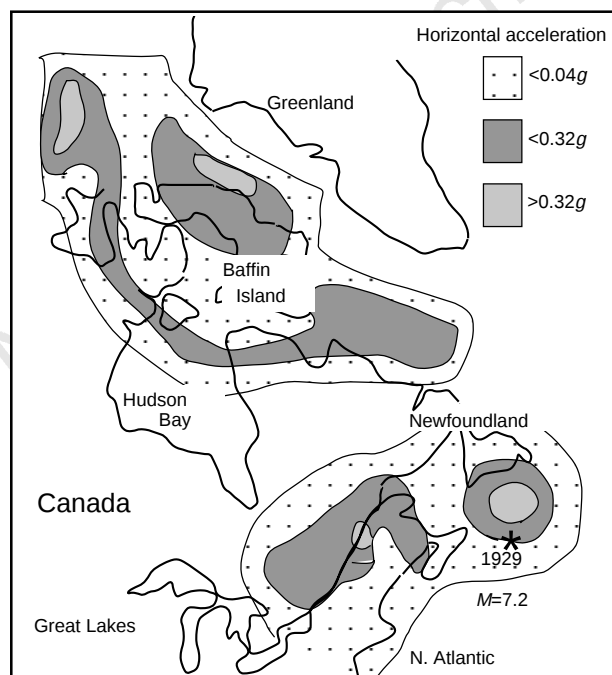


Figure 1 Map of modern seismic hazard in eastern Canada and the site of the 1929 $M=7.2$ earthquake, which destroyed structures in Newfoundland¹⁵. By comparing this map with that in Fig. 2 it can be seen that the hazard zones form offshore belts at the former edges of the Laurentide ice sheet²² and near the raised shorelines on eastern and western Baffin Island and Newfoundland¹⁸. The map is based on historical earthquake activity and shows zones in which there exists a 10% chance in 80 yr of horizontal accelerations exceeding, for example, $0.32g$ (ref. 15). In this paper we argue that similar, ice-load-induced earthquakes in these belts triggered the release of the Heinrich event ice rafts.

Numbered H1 to H6 with increasing age, Heinrich layers extend from the Labrador Sea to France and thin from west to east¹. Except for their far northern and eastern edges, significant fractions of events H1, H2, H4 and H5 are composed of carbonate materials originating from the Churchill Province in the region of northern Hudson Bay–Baffin Island²¹. The materials in events H3 and H6, in contrast, are predominantly quartz and feldspar, suggesting that these two events have sediment and ice sources different from the other four. Heinrich events also seem to coincide with vegetation changes in Florida⁴, increased ice storage in the Andes⁵, and rises in sea level recorded in the Caribbean⁸.

The main clue to the origin of Heinrich events is found by comparing total ice storage, as revealed by changes in the abundance of oxygen isotope 18 relative to isotope 16 ($\delta^{18}O$ ‰) with the aperiodicity of the time intervals between Heinrich events⁷. From 110 to 112 kyr BP forwards, the average shape of the $\delta^{18}O$ curve with time is ramp-like, with an approximately linear increase in inferred ice volume. Placement of Heinrich event times on this ramp divides its underlying area into approximately equal parts (Fig. 3).

Because the area under a linear curve grows quadratically, differences between the squares of the Heinrich event times are nearly constant. Measuring from 110 kyr BP, these differences for the first four events are within 20% of each other (Table 1). The differences for the last two events seem to be related to the large increase in the inferred rate of ice loading at the end of the glacial period⁹, particularly for the interval between H2 and H3 (Fig. 3). Adjusting for the observed 50% increase in accumulation rate in the latter interval gives differences similar to the earlier events.

The mechanical response of cratonic crust and viscous mantle to ice loading²² can account for this nonlinear regularity in the time intervals of Heinrich events (Fig. 2). The flow of the mantle from

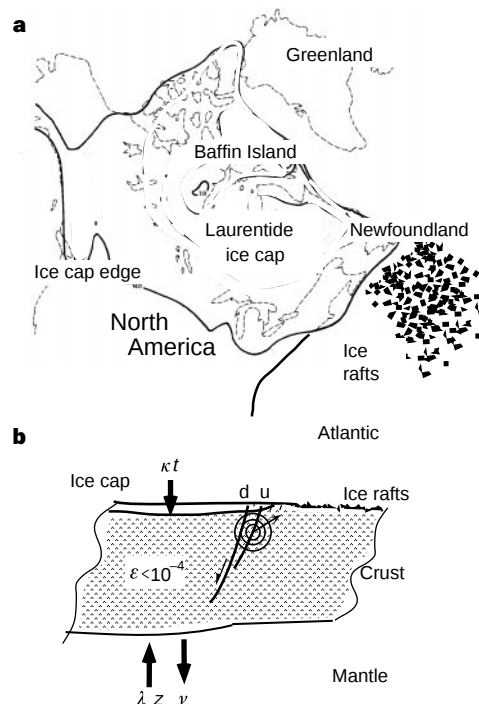


Figure 2 A simplified model of ice-load-induced earthquakes and the origin of Heinrich events. Shown in **a** is a map of the Laurentide ice sheet 14 kyr ago²², North America, and the apparent origin point of the ice rafts responsible for deposition of the Heinrich sediments¹. **b**, Cross-section through the ice sheet, ice rafts, the Atlantic and the North American crust and upper mantle. The time-dependent ice-loading function, κt , is shown by a vertical arrow. Also illustrated are the resulting mantle flow velocity, ν , the maximum crustal strain, $\epsilon < 10^{-4}$, and the isostatic reaction force, λz . The letters d and u are for the downheld and elastic rebound uplifted sides of the faults.

underneath the ice load produces crustal shear stress and strain that are concentrated at the edges of the ice sheet. Typically, the crust fails seismically at an accumulated elastic strain, ϵ , of 10^{-4} , with the size of the earthquake depending on the size of the region over which the strain has accumulated²⁰. During the strain build-up, significant aseismic deformation also takes place²⁰. Thus for every ice-load-induced elastic strain increment of 10^{-4} a major earthquake can be expected to have taken place near the edge of the ice sheet. Further, the rate of earthquake recurrence is a function of the strain rate^{20,23}, $d\epsilon/dt$. At $d\epsilon/dt = 10^{-6}\text{yr}^{-1}$, the recurrence interval of $M = 8$ earthquakes is several hundred years²³. A similar response should take place during unloading and postglacial isostatic rebound. This response includes roughly equal number of earthquakes and amount of aseismic deformation being produced as during loading, the differences being related to the strain rate.

During the ice ages the Laurentide ice sheet depressed the crust by nearly 400 m (ref. 22). Assuming that the resulting shear strain was concentrated over the equivalent of one to three thicknesses of brittle crust (30–90 km), a net strain on the order of 10^{-2} was accumulated. We hypothesize that this strain was partly accommodated by at least six major episodes of deformation that included complete failure of the brittle crust in $M = 7$ –8 earthquakes. Accommodation of 70–90% of the strain aseismically (a value often cited for tectonic earthquakes) and division of the remaining strain between six seismic episodes yields strains on the order of 10^{-4} per episode. Accumulation of the 10^{-2} strain uniformly over 10^5 years yields an average strain rate of $d\epsilon/dt = 10^{-7}\text{yr}^{-1}$, implying average recurrence intervals of several thousands of years for $M = 8$ earthquakes²³. The observed nonlinear shortening of the event intervals results from the steadily increasing ice load.

In a highly simplified version of our model (Fig. 2), the downward flow rate, $\nu = dz/dt$, of the mantle under the linearly increasing ice

load, κt , and isostatic reaction, λz , can be found from the relation

$$d\nu/dt + \nu/\tau = \kappa t - \lambda z \quad (1)$$

where τ , κ and λ are constants and z is vertical displacement. This is a simple one-dimensional dashpot/spring model of mantle flow under a linearly increasing load. When the position of the load is far from isostatic compensation (which occurs for a depression 900 m deep for an ice sheet 3,000 m thick), λz can be neglected. The relaxation time, τ , is inversely proportional to the mantle viscosity, $\tau \propto 1/\mu$. If $d\nu/dt = 0$ (that is, steady state), then a constant load results in a constant velocity, $\nu \propto 1/\mu$. With no external forcing, $\kappa = 0$ (for example, isostatic rebound after glaciation), equation (1) yields exponential relaxation of the velocity, vertical displacement, and associated strain rates, all involving the time constant τ . On the basis of rebound data, $\tau \approx 3$ kyr (refs 14, 22). Under uniformly increasing force without compensation, the flow velocity is given by

$$\nu = \kappa(\tau t - \tau^2) + (\kappa\tau^2) \exp(-t/\tau) \quad (2)$$

For $t \gg \tau$, as would be true of recurrence intervals of 10 kyr, $\nu \propto t$. Because the rate of crustal shearing is proportional to ν , then $d\epsilon/dt \propto t$. Thus, the accumulated strain in a given interval must be proportional to t^2 : $\epsilon \propto \int_{t_i}^{t_f} t dt \propto t_f^2 - t_i^2$, where t_f and t_i are the final and initial times of the interval. By analogy with tectonic earthquake recurrence intervals, we expect ice-loading earthquakes to occur at intervals separated by 10^{-4} increments in elastic strain. Under this assumption, the intervals between t_f and t_i must decrease with time. In isostatic rebound after glaciation, the earthquake strain increments remain the same, but the intervals between earthquakes increase as the strain rate decays.

Earthquakes imply faults; one may therefore ask: where are the faults responsible for the ice-loading earthquakes and why have they not been recognized as such? We expect that the greatest deformation, and hence the largest faults, occurred under the edge of the ice sheet, which was on the continental shelf (Fig. 2). To some degree the growth and flow of the ice sheet might have obliterated any associated surface rupture, making them hard to recognize directly, even if they are not underwater. Their existence, however, can be inferred from the seismicity of northeastern Canada, which includes

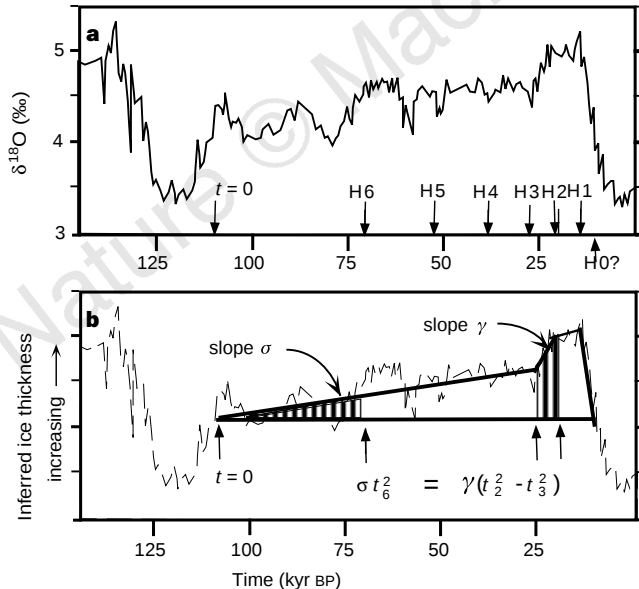


Figure 3 Evidence for increasing ice thickness in Canada during the last glaciation from measurements of oxygen isotopes. **a**, Percentage differences in $\delta^{18}\text{O}$ in benthic foraminifera as a function of time before present in thousands of years ($\delta^{18}\text{O} \text{‰}$ against time⁹); **b**, average ice thicknesses that the $\delta^{18}\text{O}$ differences imply. In **a** the times of the six Heinrich events H1 and H6 are indicated by arrows (see also Table 1; the time of the Younger Dryas event is indicated as H0). In **b** the ice thickness against time is approximated by three line segments of slopes σ , γ and σ . The segment with slope γ takes into account the 50% greater rate of ice accumulation between H3 and H2⁹. As indicated by the vertical strip pattern, starting at 110 kyr BP the Heinrich events divide the area under the line segments into equal parts. Thus, the differences in the squares of the event times stays nearly constant.

Table 1 Heinrich event and shoreline uplift intervals

Event	Date (kyr BP)		Time intervals from $t = 110$ kyr BP		
	¹⁴ C	Calendar	$(t_n - t_{n-1})$	$(t_n^2 - t_{n-1}^2)/2$	Corrected*
Heinrich ⁷					
H6	70	70	40	800	800
H5	54	54	16	768	768
H4	40	40	14	882	882
H3	28	28	10	750	750
H2	21	24	6	498	747
H1	14.5	16	8	720	720
(H0)†	–	(11.5)	(4.5)	(483)	(?)
Shoreline uplift ¹⁸					
Baffin East					
BE9	8.6		?‡		
BE8	8.		0.6		
BE7	7.5		0.5		
BE6	6.8		0.7		
BE5	6		0.8		
BE4	5		1.0		
BE3	4		1.0		
BE2	2.8		1.2		
BE1	2		0.8		
Baffin West					
BW8	6.7		?‡		
BW7	6.0		0.7		
BW6	5.8		0.2		
BW5	5.5		0.3		
BW4	5.2		0.3		
BW3	4.7		0.5		
BW2	4.2		0.5		
BW1	3.7		0.5		

* Takes into account the 50% increase in rate of iceloading between H3 and H2 (ref. 9).

† Includes the Younger Dryas event as also triggered by an ice-load-induced earthquake.

‡ Initial interval depends on origin time of uplift.

earthquakes 30 km deep (Fig. 1), and from the presence of incrementally raised shorelines (Table 1). Both the earthquakes and stranded shorelines occur near the former edges of the ice sheet.

Although only minor raised shorelines have been reported in Hudson Bay²⁴, Baffin Island seems to have eight on its western coast and nine on its eastern coast¹⁸ (Table 1). On both coasts the stranded shorelines can be correlated over distances of 70 km, have vertical separations of ~13 m, and were created between 2,000 and 9,000 years ago. Uplifted shorelines are also found in southern Newfoundland, where as many as five strands separated by an average of 12 m can be found at one location¹⁹. In this case, the strands can be correlated over tens of kilometres, but they have not been accurately dated. The erosion surfaces behind all these shorelines are not grossly tilted (a maximum of 1 m km⁻¹), suggesting that they were raised in a block-like fashion by elastic rebound instead of on a doming isostatic adjustment.

The raised shorelines lie along the edges of eastern Canada's most seismically active zone, a mostly offshore belt running from the Great Lakes to Newfoundland and Baffin Island. This belt includes the 1929 Grandbanks $M = 7.2$ earthquake on the seaward-sloping continental shelf off Newfoundland, an event that produced a tsunami that killed 27 people and cut the first transatlantic telephone cable. As Fig. 1 shows, the belt includes zones where there is roughly a 1% chance per decade of horizontal accelerations exceeding 0.3 g (ref. 15). As in 1929, such rapid movements are enough to level unreinforced buildings.

Both the raised shorelines and seismicity are consistent with our ice-loading strain model. First, the loading and unloading of the crust seems to produce roughly the same number of events: $s \times$ Heinrich events and seven or eight shoreline uplifts. Second, the strain rates and recurrence intervals scale with each other: the crust was unloaded one-and-a-half orders of magnitude faster than it was loaded. The resulting earthquake recurrence times would also be much shorter, on the order of 10^2 – 10^3 yr. Further, the intervals between the shoreline events increases with time (Table 1).

Studies of ice dynamics have led to the suggestion that the large volume of sediments associated with Heinrich events can be accounted for by a 'binge-purge' cycle in the ice cap¹². This cycle begins with ice-sheet thickening, with consequent sub-ice melting and higher seaward flow rates, which in turn produces ice-sheet thinning, slowing down and attachment of saturated subglacial regolith by freezing. The binge is estimated to take ~8 kyr and the purge ~750 yr. This mechanism has been shown to be compatible with the Heinrich event sediment volumes, but it predicts that the event intervals would be uniform and sedimentation rates slow enough to include foraminifera. Our model can account for the rapid purge by nearly instantaneous shaking and failure of the edge of the ice load—in effect, the earthquakes removed the resisting toes of the glacial ice slides. The inherent spatial variability of earthquake epicentres is also consistent with the differences in the Heinrich layer sediments and ice sources.

In addition to searching for and dating the faults implied by our model for the Heinrich events, it can also be tested by studies of sedimentary structures, relations and ages near their origins. If, for example, undisturbed Heinrich layers immediately overlie slumped or otherwise seismic disturbed sediments, our case would be stronger. Such features might be recognized and mapped with reflection profiling. The model could also be strengthened by further dating and palaeoseismic studies of the raised shorelines, as well as a complete three-dimensional numerical simulation of the coupled ice dynamics and ice-load-induced failures.

We conclude by suggesting that if it is true that the Heinrich events were driven primarily by crustal, as opposed to orbital, mechanics, then so were their associated climate variations. It may be found that some ice-age climatic variations were feedback effects produced by earthquake-related unloading of the ice sheet. Such a finding would lead to a much simplified interpretation of the timescales of ice-age climate changes. □

Received 8 May 1997; accepted 16 March 1998.

1. Heinrich, H. Origin and consequences of cyclic ice rafting in the Northeast Atlantic Ocean during the past 130,000 years. *Quat. Res.* **29**, 142–152 (1988).
2. Andrews, J. T. & Tedesco, K. Detrital carbonate-rich sediments, northwestern Labrador Sea; implications for ice-sheet dynamics and iceberg rafting (Heinrich) events in the North Atlantic. *Geology* **20**, 1087–1090 (1994).
3. Bond, G. et al. Evidence for massive discharges of icebergs into the north Atlantic ocean during the last glacial period. *Nature* **360**, 245–249 (1992).
4. Grimm, E. C., Jacobson, G. L. Jr., Watts, W. A., Hansen, B. C. S. & Maasch, K. A. A 50,000 year record of climate oscillations from Florida and its temporal correlation with the Heinrich events. *Science* **261**, 198–200 (1993).
5. Lowell, T. V. et al. Interhemispheric correlation of late Pleistocene glacial events. *Science* **269**, 1541–1549 (1995).
6. Broecker, W. S. Massive iceberg discharges as triggers for global climate change. *Nature* **372**, 431–424 (1994).
7. Bond, G. et al. Correlations between climate records from North Atlantic sediments and Greenland ice. *Nature* **365**, 143–147 (1993).
8. Baltuck, M., Dickey, J., Dixon, T. & Harrison, C. G. A. New approaches raise questions about future sea level changes. *Eos* **77**, 385–388 (1996).
9. Shackleton, N. J., Imbrie, J. & Hall, M. A. Oxygen and carbon isotope record of East Pacific Core V19–30; implications for the formation of deep water in the late Pleistocene North Atlantic. *Earth Planet. Sci. Lett.* **65**, 233–244 (1983).
10. Bond, G. & Lotti, R. Climate forcing of millennial-scale ice rafting cycles in the North Atlantic. *Eos* **75**, 332 (1994).
11. Hays, J. D., Imbrie, J. & Shackleton, N. J. Variations in the earth's orbit: pacemaker of the ice ages. *Science* **194**, 11221–1132 (1976).
12. Alley, R. B. & MacAyeal, D. R. Ice-rafted debris associated with binge-purge oscillations of the Laurentide ice sheet. *Paleoceanography* **9**, 503–511 (1994).
13. Gupta, H. K. & Chandra, R. K. (eds) *Induced Seismicity (Pure Appl. Geophys. Vol. 145, 1995)*.
14. Hasegawa, H. S. & Basham, P. W. in *Earthquakes at North-Atlantic Passive Margins: Neotectonics and Postglacial Rebound* (ed. Gregersen, S. and Basham, P. W.) 483–500 (NATO Advanced Study Institutes Series C, Vol. 266, 1989).
15. Geological Survey Canada, <http://www.seismo.nrcan.gc.ca/sheishaz4gen123.html>
16. Arvidsson, R. Fennoscandian earthquakes: whole crustal rupturing related to postglacial rebound. *Science* **274**, 744–745 (1996).
17. Johnston, A. C. A wave in the earth. *Science* **274**, 735 (1996).
18. Andrews, J. T. *A Geomorphological Study of Post-Glacial Uplift with Particular Reference to Arctic Canada* (London Institute of British Geographers, London, 1970).
19. Tucker, C. M., Leckie, D. A. & McCann, S. B. Raised shoreline phenomena and postglacial emergence in south-central Newfoundland. *Geomorph. Phys. Quat.* **36**, 165–174 (1982).
20. Lay, T. & Wallace, T. C. *Modern Global Seismology* (Academic, New York, 1995).
21. Broecker, W. *The Glacial World According to Wally* (Lamont-Doherty Earth Observatory, Palisades, 1995).
22. Cathles, L. M. III *The Viscosity of the Earth's Mantle* (Princeton Univ. Press, 1975).
23. Sieh, K. E. in *Earthquake Prediction* (ed. Simpson, D. W. and Richards, P. G.) 181–207 (American Geophysical Union, Washington, 1981).
24. Bird, J. B. *The Physiography of Arctic Canada* (Johns Hopkins Univ. Press, 1967).

Acknowledgements. We thank G. Bond and J. Park for insightful and considered reviews, and our colleagues for their patient help in understanding Heinrich events.

Correspondence and requests for materials should be addressed to P.E.M. (e-mail: pem@vaino.geo.duke.edu).

Sphenoid shortening and the evolution of modern human cranial shape

Daniel E. Lieberman

Department of Anthropology, Rutgers University, New Brunswick, New Jersey 08903-0270, USA

Crania of 'anatomically modern' *Homo sapiens* from the Holocene and Upper Pleistocene epochs differ from those of other *Homo* taxa, including Neanderthals, by only a few features. These include a globular braincase, a vertical forehead, a diminutive browridge, a canine fossa and a pronounced chin^{1–4}. Humans are also unique among mammals in lacking facial projection: the face of the adult *H. sapiens* lies almost entirely beneath the anterior cranial fossa, whereas the face in all other adult mammals, including Neanderthals, projects to some extent in front of the braincase. Here I use radiographs and computed tomography to show that many of these unique human features stem partly from a single, ontogenetically early reduction in the length of the sphenoid, the central bone of the cranial base from which the face grows forward. Sphenoid reduction, through its effects on facial projection and cranial shape, may account for the appar-

Table 1 Partial correlation matrix from cross-sectional samples of *Pan troglodytes* (top) and *Homo sapiens* (bottom)

	MFP	ACL	ASL	MFL	MCL	ECV
MFP	–	–0.292*	0.202*	0.728*	–0.124*	–0.009
ACL	–0.192*	–	0.460*	0.604*	–0.179*	0.158*
ASL	0.490*	0.211*	–	–0.424*	0.593*	–0.290*
MFL	0.663*	0.218*	–0.605*	–	0.572*	–0.316*
MCL	–0.166*	0.345*	0.572*	0.602*	–	0.753*
ECV	–0.110	–0.335*	–0.044*	–0.222	0.685*	–

* $P < 0.01$ (Fisher's r -to- z). Abbreviations. MFP, midfacial projection; ACL, anterior cranial base length; ASL, anterior sphenoid length; MFL, midfacial length; MCL, maximum cranial length; ECV, endocranial volume. See Methods for sample details and variable definitions.

ently rapid evolution of modern human cranial form, and suggests that Neanderthals and other archaic *Homo* should be excluded from *H. sapiens*.

Shortening of the sphenoid influences human cranial shape primarily by altering the spatial relationships between the face, cranial base and neurocranium in the sagittal plane, which together determine the degree of facial projection. Variations in facial projection (which is distinct from prognathism⁵) are a function of the relative anteroposterior growth of three developmentally distinct dimensions (Fig. 1a): anterior cranial base length (ACL); anterior sphenoid body length (ASL); and anteroposterior mid-facial length (MFL). ACL, between sella and the foramen caecum, elongates through bone deposition in basicranial synchondroses induced by brain growth. ASL increases as a result of forward expansion of the sphenoid body, and extends from sella to the posterior maxillary plane⁶. This plane, from the maxillary tuberosities to the junction between the middle and anterior cranial fossae, forms the boundary between the cranial base and the ethmo-maxillary complex that comprises most of the face^{6–8}. Thus ASL determines the position of the back of the face relative to the cranial base. The final dimension that influences facial projection is MFL (see Fig. 1), the minimum distance from the posterior maxillary plane to nasion. The face elongates through sutural growth along its posterior margins and from appositional growth on its anterior surfaces⁷. Histological studies show, however, that the anterior aspect of the maxilla in humans, unlike non-human primates and australopithecines, is a resorptive field^{9,10}. The human midface therefore elongates from deposition on its posterior surface, which displaces it forward relative to the posterior maxillary plane.

The contributions of ACL, ASL and MFL to facial projection were measured in recent and fossil *Homo* crania using lateral radiographs and computed tomography (CT) scans (see Methods). Longitudinal samples¹¹ indicate that these dimensions grow at different rates in *H. sapiens* (Fig. 2a). MFL elongates in a slow, skeletal growth trajectory, attaining 95% of adult length by 15–18 years, but ASL and ACL lengthen more rapidly, reaching 95% of adult size after approximately 4 and 6 years, respectively. The effects of these dimensions on facial projection were tested using partial correlation analyses of intraspecific, cross-sectional samples of humans and chimpanzees (Table 1). In these species, MFL, ACL and ASL make statistically significant and independent contributions to midfacial projection, independent of any effects of overall cranial size (maximum length and endocranial volume). MFL, ACL and ASL were all found to have

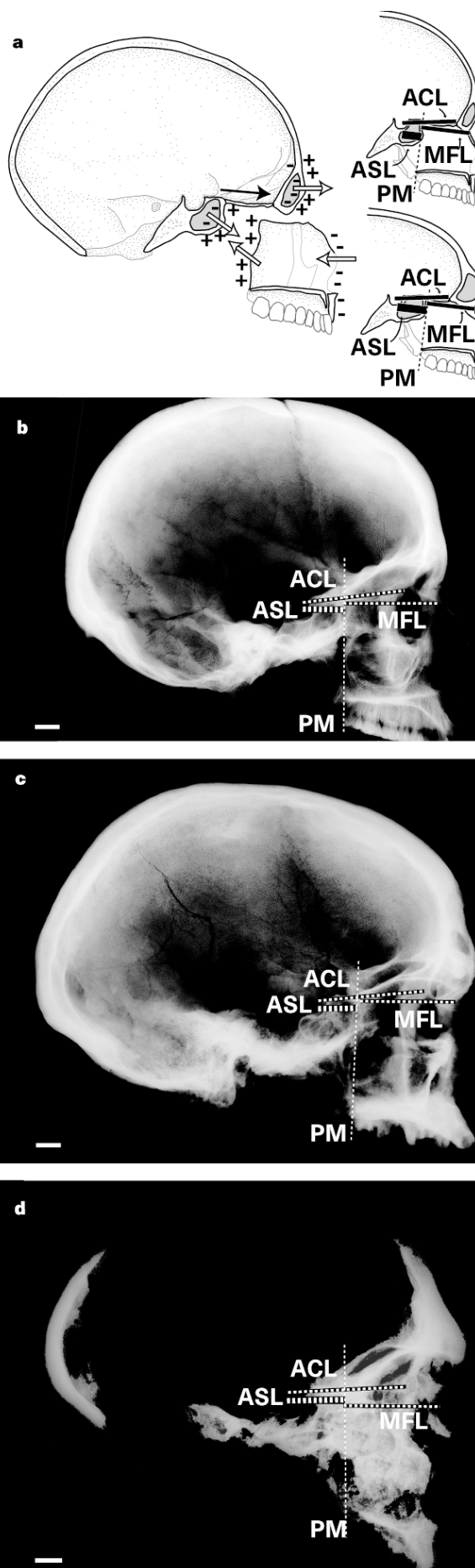


Figure 1 Cranial growth in modern and archaic *Homo*. **a**, Left, midsagittal diagram of anteroposterior growth processes in cranial base, braincase and face. Black arrows indicate synchondrosal growth; white arrows indicate appositional growth (+) and resorption (–). Right, diagrams of *H. sapiens* (top) and Neanderthal (bottom) show the posterior maxillary (PM) plane, anterior cranial base length (ACL), midfacial length (MFL) and anterior sphenoid length (ASL) (see Methods). **b–d**, Lateral radiographs of: **b**, recent *H. sapiens* (Egyptian male); **c**, male Upper Pleistocene *H. sapiens* (Cro-Magnon I); **d**, female Neanderthal (Gibraltar I). ASL in the female Neanderthal is 20% longer than in the Upper Pleistocene male; ACL and MFL are approximately 15% shorter. Scale bars, 10 cm.

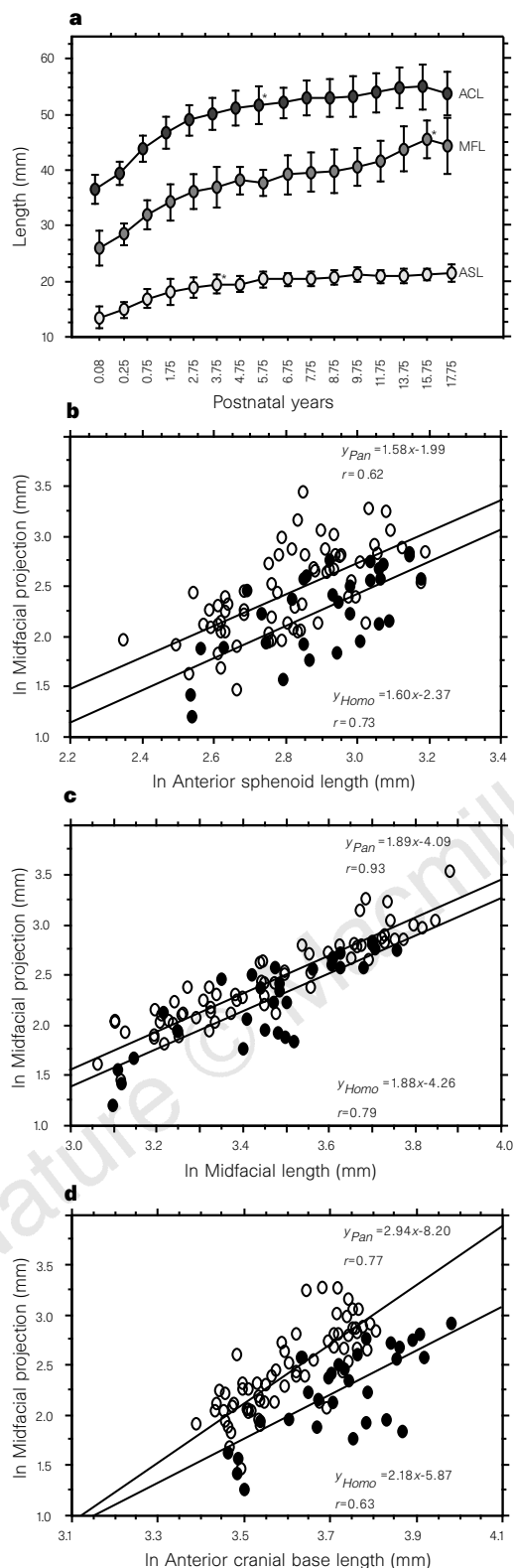


Figure 2 Ontogeny of facial projection in *Homo sapiens* and *Pan troglodytes*. **a**, Growth rates in longitudinal sample of *H. sapiens*¹⁶ of anterior cranial base length (ACL), midfacial length (MFL) and anterior sphenoid length (ASL). Bars represent 1 standard deviation; asterisks indicate 95% of mean adult size. **b–d**, Log-transformed least-squares regression of these dimensions with midfacial projection in cross-sectional samples of *H. sapiens* (filled circles) and *P. troglodytes* (open circles). See Methods for measurement and sample details.

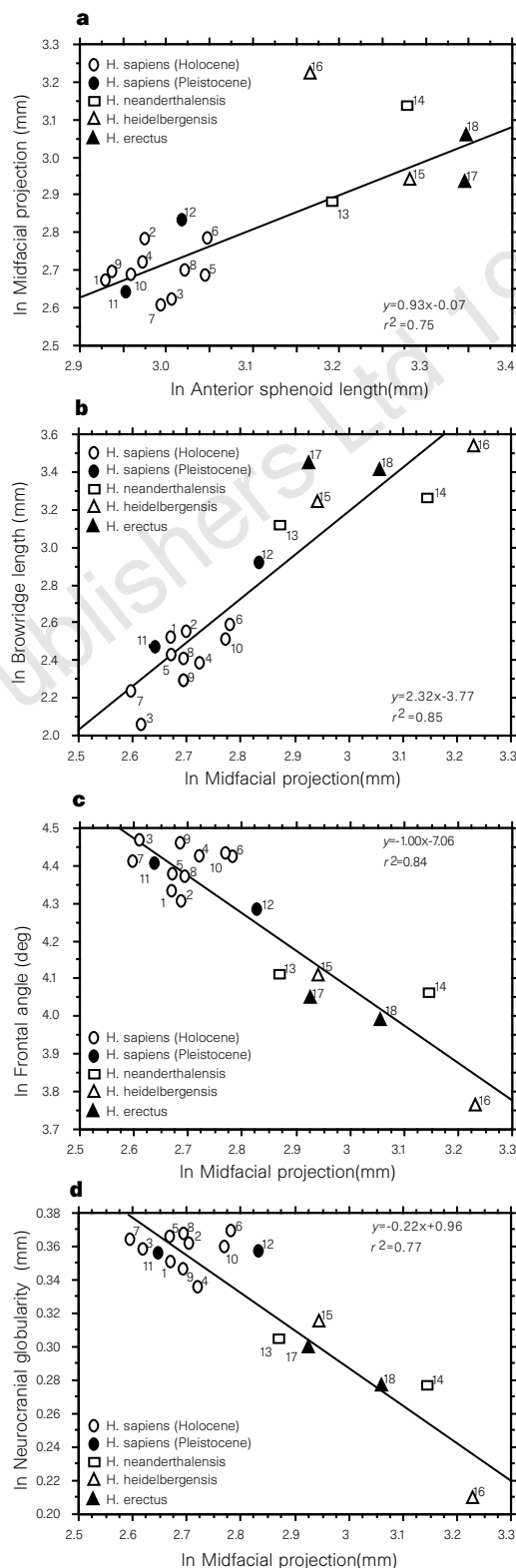


Figure 3 Effects of anterior sphenoid length (ASL) on facial projection (MFP) and other aspects of cranial shape in *Homo*. **a**, Log-transformed least-squares regression of ASL with MFP (including *H. heidelbergensis*, $y = 1.01x - 0.30$; $r^2 = 0.620$). **b–d**, Log-transformed least-squares regressions between MFP and supraorbital length, frontal angle and midsagittal cranial curvature (see Methods). Samples include: 1, female Australians; 2, male Australians; 3, female Chinese; 4, male Chinese; 5, female Italians; 6, male Italians; 7, female Egyptians; 8, male Egyptians; 9, female Ashanti; 10, male Ashanti; 11, female Pleistocene *H. sapiens*; 13, female Neanderthal; 14, male Neanderthals; 15, female *H. heidelbergensis*; 16, male *H. heidelbergensis*; 17, female *H. erectus*; and 18, male *H. erectus*.

Table 2 Intertaxon analysis of variance of facial, basicranial and neurocranial dimensions and spatial relationships in *Homo*.

Taxon	N (m/f)	ASL (mm)	ACL (mm)	MFL (mm)	LFL (mm)	MFP (mm)	FRA (mm)	SOL (mm)	GLO (mm)	MCL (mm)	ECV (cm ³)
<i>H. sapiens</i> (Holocene)	100 (50/50)	19.9 (2.0)	47.8 (2.9)	40.6 (3.6)	45.0 (3.8)	14.6 (2.4)	81.4 (6.4)	11.1 (2.7)	1.41 (0.1)	177.7 (8.2)	1,368.4 (149.4)
<i>H. sapiens</i> (Pleistocene)	6 (4/2)	20.0 (1.8)	49.6 (3.7)	45.4* (2.3)	50.8* (6.5)	16.1 (2.1)	77.5 (5.4)	17.0* (3.0)	1.43 (0.1)	193.0* (10.0)	1,478.8 (114.8)
<i>H. neanderthalensis</i>	5 (3/2)	25.9† (2.9)	52.6* (7.1)	46.9* (4.9)	52.6* (7.5)	21.4*† (4.8)	59.0*† (6.0)	24.8*† (2.8)	1.33† (0.1)	198.3* (11.3)	1,447.5 (201.1)
<i>H. heidelbergensis</i>	3 (2/1)	24.8*† (5.1)	48.8 (9.1)	45.3* (8.5)	53.0* (9.5)	24.6*† (10.1)	49.0*† (10.8)	31.5*† (6.0)	1.32† (0.0)	200.3* (13.5)	1,201.7* (93.9)
<i>H. erectus</i>	2 (1/1)	28.4*† (0.1)	35.3* (1.6)	39.6 (3.4)	48.7	20.0*† (1.8)	55.5*† (2.1)	30.5*† (0.6)	1.34*† (0.0)	190.5* (20.5)	935.5*† (185.9)

See Methods for sample details and variable definitions. Standard deviations are in parentheses. Abbreviations: ASL, anterior sphenoid length; ACL, anterior cranial base length; MFL, midfacial length; LFL, lower facial length; MFP, midfacial projection; FRA, frontal angle; SOL, supraorbital length; GLO, neurocranial curvature (globularity); MCL, maximum cranial length; ECV, endocranial volume.

* Significantly different ($P < 0.01$, Scheffé's F) from Holocene *H. sapiens*.

† Significantly different ($P < 0.01$, Scheffé's F) from Pleistocene *H. sapiens*.

strong, positive allometric effects on facial projection (Fig. 2b–d). Variations in facial projection in chimpanzees, humans and presumably fossil hominids derive, therefore, from differences in MFL, ACL and ASL, all of which influence the position of the front of the face relative to the cranial base and neurocranium.

Although facial, cranial base and sphenoid length each affect facial projection in modern humans and other primates, comparisons between taxa show that variations in ASL provide the main structural basis for the differences in facial projection between Pleistocene 'anatomically modern' *H. sapiens* and archaic *Homo*. Midfacial length, lower facial length, anterior cranial base length and endocranial volume do not differ significantly between adult Pleistocene modern *H. sapiens*, Neanderthals and *H. heidelbergensis* (Table 2). In contrast, ASL is roughly 30% shorter ($P < 0.001$) in Holocene and Pleistocene modern *H. sapiens* than in Neanderthals and other archaic *Homo* taxa (Fig. 1b–d). Anterior sphenoid shortening reduces facial projection in crania of Pleistocene and Holocene *H. sapiens* compared with crania of archaic *Homo* by positioning the posterior maxillary plane, and hence the posterior margin of the face, closer to the middle cranial fossa. ASL accounts for approximately 75% of the variation in facial projection among Pleistocene fossil and Holocene taxa of the genus *Homo* (Fig. 3a; excluding *H. heidelbergensis* males, in which a significantly longer MFL than other archaic *Homo* taxa ($P < 0.05$) contributes to an extreme degree of facial projection).

Reduced facial projection in modern humans, which results primarily from sphenoid shortening, results in other phylogenetically and functionally significant differences in craniofacial shape. Facial projection is the main influence on browridge size and frontal angulation in non-human primates^{12–14}, and contributes to roughly 85% of the variation in these features in *Homo* (Fig. 3b, c). Reduced facial projection also increases overall cranial globularity in modern humans by decreasing cranial length relative to endocranial volume (Fig. 3d). Finally, decreased facial projection positions the maxilla closer to the foramen magnum and the temporo-mandibular joint in *H. sapiens* than in archaic *Homo*, reducing the length of the oropharynx and the load arm of the chewing muscles^{15,16}. But sphenoid shortening is not the only source of variation in facial projection and its consequent effects in *Homo*. Within *Homo*, r^2 is 0.74 between ASL and browridge length, 0.65 between ASL and frontal angle, and 0.47 between ASL and cranial globularity. Within *H. sapiens*, for example, variations in facial projection result mostly from differences in MFL, which is approximately 12% shorter in Holocene than Pleistocene populations ($P < 0.01$), and from a similar decrease in overall cranial size (Table 2). These factors account for the considerable morphological variability in browridge size and other features in Pleistocene *H. sapiens* crania, such as that evident in the Skhul and Qafzeh hominids¹⁷.

Given the close phylogenetic relationship between *H. sapiens*, Neanderthals and other Pleistocene hominid taxa, it should perhaps

not be surprising that many unique features of the modern human cranium apparently result from a single, ontogenetically early, shift in cranial base growth. The discrete, derived developmental basis of sphenoid shortening in *H. sapiens*, together with its effects on craniofacial shape, may explain how modern humans might have evolved rapidly from more archaic forms, possibly as a distinct clade, and arguably as a distinct phylogenetic species^{18–20}. Other aspects of cranial shape that distinguish recent and Pleistocene modern *H. sapiens* populations^{4,21} probably result from decreases in facial length, brain size and overall skeletal robusticity^{4,22,23}. The complex, integrated nature of craniofacial growth and function make it unlikely, however, that any single selective advantage can account for the evolution of sphenoid shortening and the unique craniofacial configuration of *H. sapiens*. One possibility is that a shorter sphenoid, by decreasing the oropharynx length, is an adaptation for speech. If the hyoid and larynx had the same low position relative to the mandible and cranial base in archaic *Homo* that they do in *H. sapiens*²⁴, a shorter sphenoid would contribute to the unique proportions of the human vocal tract, in which the horizontal component is roughly equal in length to its vertical component, rather than markedly longer as in other primates²⁵. This configuration improves the ability to produce acoustically distinct speech sounds^{25–27}, and so may have provided some advantage for sphenoid reduction and its consequent effects on facial projection in *H. sapiens*. □

Methods

Samples used. Lateral radiographs were taken of 20 adult (10 males and 10 females) recent *H. sapiens* crania from Australia, south China, Europe (Italy), north Africa (Egypt), and sub-Saharan Africa (Ashanti) from the American Museum of Natural History (AMNH); a pooled-sex cross-sectional sample ($n = 30$) of modern human crania from the Cleveland Museum of Natural History (CMNH); and a longitudinal sample of 16 males and 16 females from the Denver Growth Study radiographed from 1 month to 18 years after birth¹¹. A cross-sectional, pooled-sex sample ($n = 69$) of *Pan troglodytes* (chimpanzee) was radiographed from collections at the AMNH, the CMNH and the Museum of Comparative Zoology (Harvard). The cross-sectional human and chimpanzee samples include equal numbers of individuals from four dental stages (before eruption of the first upper molar (M^1); after M^1 eruption before M^2 eruption; after M^2 eruption but before M^3 eruption; and after M^3 eruption).

The fossil sample includes crania for which radiographs and/or CT scans are available and which are sufficiently complete to reconstruct the posterior maxillary plane and other aspects of facial position relative to the basicranium and neurocranium. The posterior maxillary plane maintains an angle of 90° with the neutral horizontal axis of the orbit throughout postnatal development in *H. sapiens* (angle $x = 89.9^\circ$, s.d. = 1.7, $n = 353$) as it does in all mammals, including *P. troglodytes*^{6,8}. This constant angle allows accurate reconstruction of the posterior maxillary plane in fossils well-preserved maxillary tuberosities and orbits. Fossils used were split by sex into the following groups: male early modern *H. sapiens* (Cro Magnon I, Oberkassel I, Skhul IV and V); female early

modern *H. sapiens* (Abri Pataud, Obercassel II); male Neanderthals (La Chapelle aux Saints, La Ferrassie I, Monte Circeo, La Quina V); female Neanderthals (Gibraltar I); male *H. heidelbergensis* (Broken Hill, Petralona); female *H. heidelbergensis* (Steinheim); male *H. erectus* (OH 9); female *H. erectus* (KNMR-ER 3733). I radiographed all crania except for Skhul IV (B. Arensburg), Petrolona (C. Stringer), KNM-ER 3733 (A. Walker) and Obercassel I and II, Monte Circeo, La Quina V (T. Molleson). F. Spoor provided CT scans of the OH 9, Broken Hill and Steinheim specimens.

Measurements. Linear and angular measurements were taken from traced radiographs using digital calipers accurate to 0.01 mm, and a protractor accurate to 1°. Measurements include: ASL (anterior sphenoid body length), the minimum distance from the sella to the posterior maxillary plane; ACL (anterior cranial base length), from the sella to the foramen caecum; MFL (midfacial length), the minimum distance from the posterior maxillary plane to nasion; LFL (lower facial length), from the anterior nasal spine to the posterior nasal spine; MFP (midfacial projection) from nasion to the foramen caecum (perpendicular to the posterior maxillary plane); FRA (frontal angle) from the metopion to the base of the frontal squama relative to the Frankfurt horizontal; SOL (supraorbital length) from the glabella to fronton (perpendicular to the posterior maxillary plane); GLO (neurocranial curvature or globularity) from the glabella to the opistocranium; and ECV (endocranial volume), which was measured by filling crania with beads; estimates of fossil endocranial volume are from ref. 28. For landmark definitions, see ref. 29.

Received 6 October 1997; accepted 10 February 1998.

- Day, M. H. & Stringer, C. B. in *L'Homo erectus et la Place de l'Homme de Tautavel parmi les Hominides Fossiles* Vol. 2 (ed. De Lumley, H.) 814–846 (Louis-Jean, Nice, 1982).
- Lieberman, D. E. Testing hypotheses about recent human evolution from skulls: integrating morphology, function, development and phylogeny. *Curr. Anthropol.* **36**, 159–197 (1995).
- Kidder, J. H., Jantz, R. L. & Smith, F. H. in *Continuity or Replacement: Controversies in Homo sapiens Evolution* (eds Bräuer, G. & Smith, F. H.) 157–177 (A.A. Balkema, Rotterdam, 1992).
- Lahr, M. M. *The Evolution of Modern Human Cranial Diversity* (Cambridge Univ. Press, 1996).
- Bilsborough, A. & Wood, B. A. Cranial morphometry of early hominids: facial region. *Am. J. Phys. Anthropol.* **76**, 61–86 (1988).
- Enlow, D. H. & Azuma, M. in *Morphogenesis and Malformations of the Face and Brain* (ed. Langman, J.) 217–230 (Harper and Row, New York, 1975).
- Enlow, D. H. *Facial Growth* 3rd edn (Saunders, Philadelphia, 1990).
- Bromage, T. G. The ontogeny of *Pan troglodytes* craniofacial architectural relationships and implications for early hominids. *J. Hum. Evol.* **23**, 235–251 (1992).
- Duterloo, H. S. & Enlow, D. H. A comparative study of cranial growth in *Homo* and *Macaca*. *Am. J. Anat.* **127**, 357–368 (1970).
- Bromage, T. G. Ontogeny of the early hominid face. *J. Hum. Evol.* **18**, 751–773 (1989).
- MacCammon, R. *Human Growth and Development* (C. C. Thomas, Springfield, 1970).
- Weidenreich, F. The brain and its rôle in the phylogenetic transformation of the human skull. *Trans. Am. Phil. Soc.* **31**, 321–442 (1941).
- Shea, B. T. On aspects of skull form in African apes and orangutans, with implications for hominid evolution. *Am. J. Phys. Anthropol.* **68**, 329–342 (1985).
- Ravosa, M. J. Ontogenetic perspective on mechanical and nonmechanical models of primate circumorbital morphology. *Am. J. Phys. Anthropol.* **85**, 95–112 (1991).
- Trinkaus, E. The Neandertal face: evolutionary and functional perspectives on a recent hominid face. *J. Hum. Evol.* **16**, 429–443 (1987).
- Spencer, M. A. & Demes, B. Biomechanical analysis of masticatory system configuration in Neanderthals and Inuits. *Am. J. Phys. Anthropol.* **91**, 1–20 (1995).
- Corruccini, R. M. Metrical reconsideration of the Skhul IV and IX and Border Cave 1 crania in the context of modern human origins. *Am. J. Phys. Anthropol.* **87**, 433–445 (1992).
- Krings, M. et al. Neandertal DNA sequences and the origin of modern humans. *Cell* **90**, 19–30 (1997).
- Stringer, C. B. & Andrews, P. A. Genetic and fossil evidence for the origin of modern humans. *Science* **239**, 1263–1268 (1988).
- Schwartz, J. H. & Tattersall, I. Significance of some previously unrecognized apomorphies in the nasal region of *Homo neanderthalensis*. *Proc. Natl Acad. Sci. USA* **93**, 10852–10854 (1996).
- Howells, W. W. *Skull Shapes and the Map* (Peabody Museum Papers no. 79, Cambridge, 1989).
- Lieberman, D. E. How and why recent humans grow thin skulls: experimental data on systemic cortical robusticity. *Am. J. Phys. Anthropol.* **101**, 217–236 (1996).
- Ruff, C. B., Trinkaus, E. & Holliday, T. W. Body mass and encephalization in Pleistocene *Homo*. *Nature* **387**, 173–176 (1997).
- Arensburg, B., Schepartz, L. A., Tillier, A. M., Vandermeersch, B. & Rak, Y. A reappraisal of the anatomical basis for speech in Middle Paleolithic hominids. *Am. J. Phys. Anthropol.* **83**, 137–146 (1990).
- Stevens, K. N. in *Human Communication: A Unified View* (eds David, E. E. & Denes, P. B.) 51–66 (McGraw Hill, New York, 1972).
- Fant, G. *Acoustic Theory of Speech Production* (Moulton, The Hague, 1960).
- Lieberman, P. *The Biology and Evolution of Language* (Harvard Univ. Press, Cambridge, MA, 1984).
- Aiello, L. & Dean, C. *An Introduction to Human Evolutionary Anatomy* (Academic, London, 1990).
- White, T. D. & Folkens, P. A. *Human Osteology* (Academic, San Diego, 1991).

Acknowledgements. I thank B. Arensburg, M. Chech, A. W. Crompton, J.-J. Hublin, L. Humphrey, L. Jellema, R. Kruszinski, P. Langane, B. Latimer, T. Molleson, D. Pilbeam, A. Poole, C. Ross, G. Sawyer, F. Spoor, C. Stringer, I. Tattersall and A. Walker for help and/or permission to obtain radiographs and CT scans; R. McCarthy, K. Mowbray and R. Bernstein for help with collecting data; and L. Aiello, F. A. Jenkins Jr, P. O'Higgins, O. Pearson, D. Pilbeam, G. P. Rightmire, N. Shubin, R. J. Smith, I. Tattersall and B. Wood for comments on the manuscript. This work was supported by Rutgers University.

Correspondence and requests for materials should be addressed to the author (e-mail: danlieb@ci.rutgers.edu).

Gene transfer to the nucleus and the evolution of chloroplasts

William Martin*, Bettina Stoebe†, Vadim Goremykin*, Sabine Hansmann*, Masami Hasegawa‡ & Klaus V. Kowallik†

* Institut für Genetik, Technische Universität Braunschweig, Spielmannstrasse 7, 38023 Braunschweig, Germany

† Botanisches Institut, Heinrich-Heine-Universität Düsseldorf, Universitätsstrasse 1, 40225 Düsseldorf, Germany

‡ The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106, Japan

Photosynthetic eukaryotes, particularly unicellular forms, possess a fossil record that is either wrought with gaps or difficult to interpret, or both. Attempts to reconstruct their evolution have focused on plastid phylogeny, but were limited by the amount and type of phylogenetic information contained within single genes^{1–5}. Among the 210 different protein-coding genes contained in the completely sequenced chloroplast genomes from a glaucocystophyte, a rhodophyte, a diatom, a euglenophyte and five land plants, we have now identified the set of 45 common to each and to a cyanobacterial outgroup genome. Phylogenetic inference with an alignment of 11,039 amino-acid positions per genome indicates that this information is sufficient—but just barely so—to

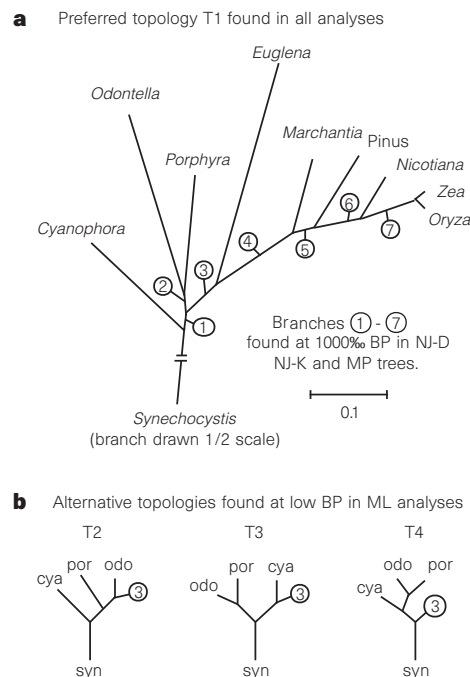


Figure 1 Plastid phylogeny interpreted from chloroplast proteins. **a**, Rooted nine species neighbour-joining (NJ) tree of Dayhoff distances for 11,039 amino-acid positions from 45 orthologous proteins common to these chloroplast genomes and *Synechocystis*. All seven branches of this topology (T1) are found in 1,000/1,000 bootstrap samples in maximum parsimony (PROTPARS or PHYLP) and NJ analysis using either Kimura or Dayhoff distances. The root of the tree is assumed from the model that all plastids sampled here arose from a common chloroplast ancestor. Branches are numbered 1–7 for convenience (see text). The scale bar indicates Dayhoff distance. **b**, Alternative topologies T2, T3 and T4 detected in protein maximum likelihood¹⁰ analyses using the JTT-F model. Taxon abbreviations are given in Methods; branch 3 is the same as in **a**.

identify the rooted nine-taxon topology. We mapped the process of gene loss from chloroplast genomes across the inferred tree and found that, surprisingly, independent parallel gene losses in multiple lineages outnumber phylogenetically unique losses by more than 4:1. We identified homologues of 44 different plastid-encoded proteins as functional nuclear genes of chloroplast origin, providing evidence for endosymbiotic gene transfer to the nucleus in plants.

By using literature and database searches, we recorded the presence or absence of the 210 protein-coding genes in nine chloroplast genomes (that is, all known functional protein-coding genes and those designated *ycf5*⁶). We identified 46 protein-coding genes whose sequences are known for all nine genomes. Homologues of these from the *Synechocystis* PCC6803 genome⁷ were also retrieved. *RbcL* was excluded from further analysis because some *rbcL* genes in this taxon sample are related by duplication⁸ (paralogy), leaving 45 proteins in the data set. Individual ten-species alignments were constructed for each gene and concatenated to produce a data set consisting of 11,039 amino-acid positions per genome. Trees constructed from these data using the neighbour-joining method⁹, with distances estimated using either the Dayhoff matrix or Kimura's method (see Methods) in addition to parsimony analysis, revealed only one topology (T1 in Fig. 1): that is, all seven branches were found in all 1,000 bootstrap samples by all three methods.

T1 is also strongly preferred in protein maximum likelihood (ML)¹⁰ analysis of the concatenated data using the JTT-F model, being found at a frequency of 0.8238 in 10,000 bootstrap samples. But ML detected alternative topologies that the other methods did not: T2 and T3 (Fig. 1) were found with frequencies of 0.0607 and 0.1155, respectively. To investigate alternative topologies further, we used ML to compare the concatenate data analysis to results obtained from individual proteins (Table 1). In individual analyses, only 11 of the 45 proteins prefer T1 over the alternative topologies at $P = 0.95$: *petB*, *psaB*, *psbC*, *psbE*, *psbI*, *psbJ*, *rpl2*, *rpl14*, *rpl16*, *rps3*, *rps7*. T1 is rejected with a difference in log-likelihood larger than twice its standard error by only four proteins: *rps8* and the subunits of the chloroplast DNA (cpDNA) encoded RNA polymerase *rpoB*, *rpoC1* and *rpoC2* (see below). The remaining 30 proteins do not support T1 individually but neither do they reject it at $P = 0.95$. No evidence for widespread lateral gene transfer was detected. Lack of support for T1 by single proteins may be due to sampling error as a result of the small numbers of sites contained within the individual alignments. When the 1,081 sites containing insertions or deletions are excluded from the analysis, bootstrap proportions (BPs) for T1, T3 and T4 undergo only very minor changes (Table 1), indicating that such sites have only a small influence on these data.

The BP for T1 increased from 0.824 (concatenate) to 0.925 across separate analyses (Table 1), raising the question of whether the single tree of concatenated sequences or the combined result of the 45 individual trees is more reliable. To investigate this, we used the

Akaike information criterion (AIC)¹¹, where AIC is defined as $-2\ln L + 2N$, with $\ln L$ being log-likelihood and N the number of parameters. A model that minimizes the AIC is preferable¹¹. The JTT-F model uses the amino-acid frequencies of the data (degree of freedom is 19) and, for a tree with ten operational taxonomical units (OTUs), 17 branch lengths must be estimated. Therefore, the number of parameters is 36 ($= 17 + 19$) for each ML analysis. For tree 1 and the concatenated analysis including gap sites in Table 1, AIC was $2 \times 124,517.7 + 2 \times 36 = 249,107.4$. For T1 from separate analyses, AIC was $2 \times 118,401.8 + 2 \times 36 \times 45 = 240,043.6$. Thus, AIC for the separate analyses is lower than that for the concatenated analysis. This indicates that the separate analyses approximate the underlying evolutionary process better than the concatenated analysis, which assumes homogeneity of substitution process across genes. This, in turn, indicates that the combined result of separate analyses should be more reliable than the one obtained from the concatenated data. Therefore, the higher BP of T1 obtained across separate analyses compared with concatenated analysis provides further support for T1. All three subunits of the cpDNA-encoded RNA polymerase (*rpoB*, *rpoC1* and *rpoC2*) strongly reject T1 and support T2, indicating that *Odontella's rpo* genes tend to branch with their homologues from the chlorophyte lineage. This could reflect paralogy, stochastic or other¹² factors that may have differentially affected the evolution of *rpo* genes in the *Odontella/Porphyra* lineages. However, when the *rpo* genes are excluded from the analysis, the BP for T1 does not increase; rather, it decreases to 0.72 (Table 1), probably because the *rpo* genes reject T3 and T4 even more strongly than T1.

We observed differences in amino-acid composition across genomes: A + T-rich genomes (that is, *Euglena*) tended to possess higher frequencies of A + T-rich codons, whereas the converse was observed for *Synechocystis* (G+C-rich). This might influence the substitution process at sites under low constraint. We therefore used a variant of ML taking into account the suggestion¹³ that amino acids which frequently substitute may influence topology. We carried out ML analysis after converting all valine residues into isoleucines and all lysine residues into arginines ($V = I$, $K = R$) or including leucine and methionine into the first group ($I = V = L = M$, $R = K$), but did not obtain significant differences from the analysis of the original data in either case.

Therefore, ML strongly favours T1 in all types of analysis. Furthermore, support for T1 increases as the amount of data analysed increases (Table 1). Although ML, like all methods of phylogenetic inference, can fail if the assumed substitution model is incorrect^{14–16}, the other methods do not detect topologies other than T1 at all. The most straightforward interpretation of these findings is that chloroplast genomes (just barely) contain enough information to infer fully the properly rooted phylogeny of these nine photosynthetic organelles.

Chloroplast proteins resoundingly support branches 4–7 in Fig. 1—branches that are known to be correct from the fossil record—

Table 1 Topologies of cpDNA-encoded proteins detected by maximum likelihood

Table 1. Topologies of opDNA-encoded proteins detected by maximum likelihood											
No. of genes	Type of sites	No. of sites	Type of analysis	T1		T2		T3		T4	
				lnL	BP*	Δ lnL $\pm$ s.e.	BP	Δ lnL $\pm$ s.e.	BP	Δ lnL $\pm$ s.e.	BP
45	All	11,039	Separate	-118,401	9,246	177.9 $\pm$ 48.4†	1	47.1 $\pm$ 43.4	647	65.2 $\pm$ 41.7	106
			Concatenate	-124,517	8,238	64.1 $\pm$ 39.7	607	36.9 $\pm$ 35.0	1,155	99.2 $\pm$ 25.6†	0
	Excl. gaps	9,958	Separate	-103,207	9,030	194.0 $\pm$ 48.2†	0	41.0 $\pm$ 42.3	776	54.2 $\pm$ 41.0	194
			Concatenate	-107,972	8,367	109.8 $\pm$ 38.7†	21	29.4 $\pm$ 28.6	1,606	71.5 $\pm$ 25.1†	6
42‡	All	7,864	Separate	-72,762	7,199	247.4 $\pm$ 48.4†	0	32.2 $\pm$ 35.4	730	19.7 $\pm$ 35.5	2,071
			Concatenate	-76,328	6,880	232.7 $\pm$ 36.81†	0	23.7 $\pm$ 23.1	1,623	23.2 $\pm$ 23.0	1,497
	Excl. gaps	7,660	Separate	-71,010	7,670	252.0 $\pm$ 48.1†	0	33.3 $\pm$ 35.6	736	24.2 $\pm$ 35.6	1,594
			Concatenate	-74,531	7,451	236.3 $\pm$ 36.8†	0	22.8 $\pm$ 23.4	1,365	26.1 $\pm$ 23.0	1,184

$\Delta \ln L \pm \text{s.e.}$ indicates the difference in log likelihood relative to T1 (the preferred topology in each case) and one standard error of that difference. The TOTALML program in MOLPHY¹⁰ was used to evaluate the total evidence of separate analyses of individual proteins.

* Bootstrap proportion for topology in 10,000 replicates $\times 10^4$ using the REML method¹⁰.

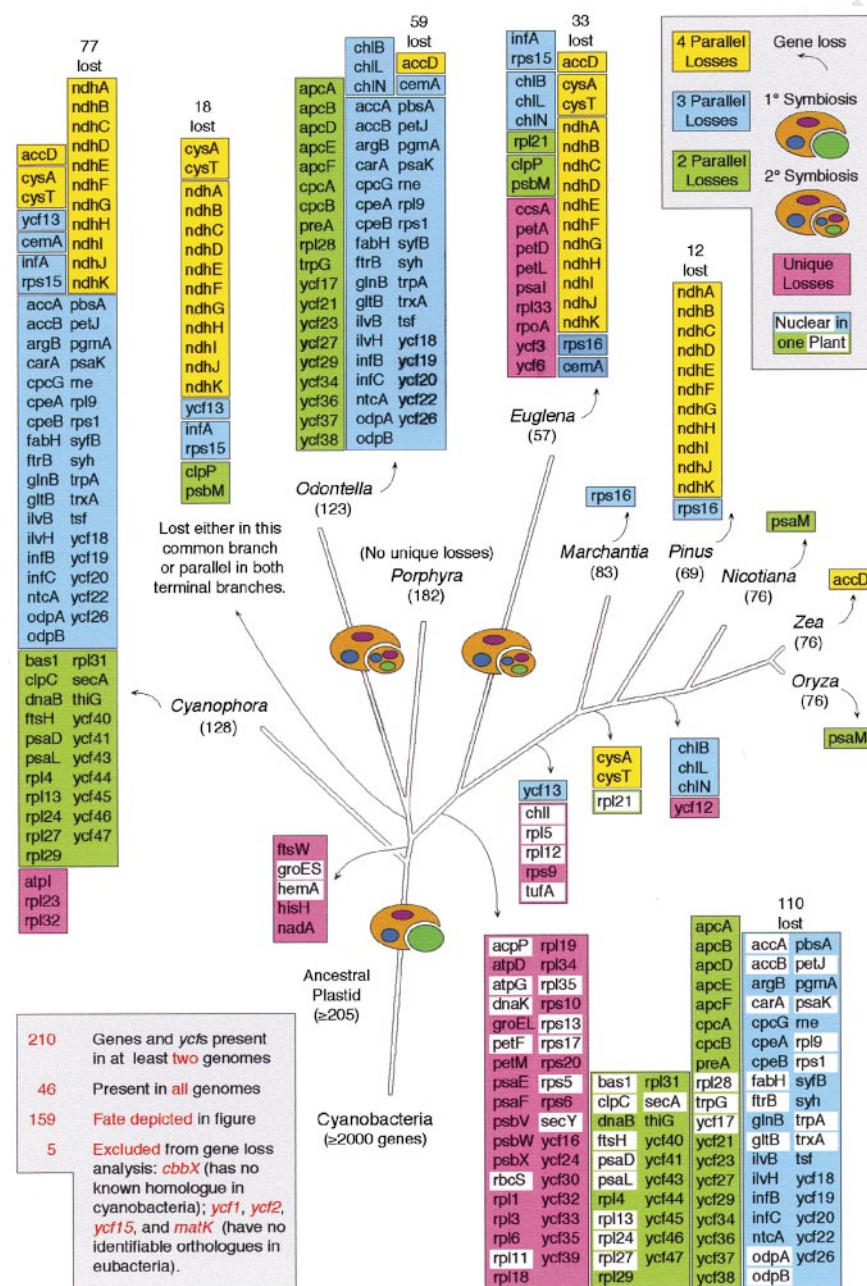
† Significantly worse than T1 ($\Delta \ln L > 2 \text{ s.e.}$).

‡ Excluding *rpoB*, *rpoC1* and *rpoC2*, which favour T2 and reject T1 at $P = 0.95$ (see text).

and clarify several unresolved issues in the early evolution of photosynthetic organelles. First, branch 3 confirms the chlorophytic ancestry of *Euglena*'s three-membrane bound plastids, as first suggested by Gibbs¹⁷, who argued that *Euglena*'s plastids arose through secondary symbiosis by capture of a eukaryotic chlorophyte in the kinetoplastid lineage (trypanosomes and relatives), with loss of the chlorophyte nucleus. Common ancestry of the euglenoid and trypanosomatid nucleocytoplasmic lineages has never been seriously challenged¹⁸, and is borne out by 18S ribosomal RNA and protein data¹⁹. *Euglena*'s plastid ancestry, however, has been unrecoverable with rRNA data. Most analyses grouped its plastids with rhodophytes and only a few detected branching with chlorophytes, with minimal statistical support^{3,4,15}. In the case of rRNA, this is partly due to base-composition bias¹⁵, but is more generally attributable to the limited information contained in any individual gene. Our data also strongly support a common ancestry for the diatom (*Odontella*) and rhodophyte (*Porphyra*) plastid genomes. In the light of molecular and cytological evidence indicating a common descent of the diatom nuclear lineage with oomycetes²⁰, our findings reflect a secondary symbiotic origin of

Odontella's four membrane-bound plastids²¹. The chlorophyll *a* and phycobilin-containing cyanelles of the glaucocystophyte *Cyanophora paradoxa*, which have retained not only phycobilisomes but also a eubacterial peptidoglycan wall^{3,22}, branch basal to the chlorophyte and rhodophyte plastids, a position that has been impossible to establish with confidence using data from single genes^{3,4}. The data analysed here provide strong support for this basal position.

We have mapped onto the topology T1 of Fig. 1 all 205 genes demonstrated to exist in the cpDNA molecule ancestral to the genomes surveyed (Fig. 2). Gene-loss events that are unique to individual lineages account for the fate of only 58 genes across the tree. The majority of genes (101) have undergone parallel loss in independent lineages. In a survey of only nine genomes, we found that 44 genes were lost twice independently, 43 genes have undergone three parallel losses, and 14 genes have been lost four times in independent lineages. The numbers of unique (6) and parallel (40) events indicated in Fig. 2 reflect the unlikely but event-minimizing premise that loss occurred in the form of the indicated gene blocks. The visible, ongoing process of *ndh* gene loss from *Pinus* cpDNA²³



indicates that genes are lost individually, rather than as blocks; the true numbers of parallel events are therefore probably much greater. If we use a more realistic model and count each loss of an individual gene as one event, the ratio of parallel to unique events becomes 4.7 (58 unique versus 273 parallel), an overwhelming excess of parallel losses.

Some of these genes have been lost altogether, for example constituents of phycobilisomes (*apc*, *cpc* and *cpe* gene products) in the chlorophyte and *Odontella* lineages, whereas others have been transferred to the nucleus. We have documented 44 bona fide cases of functional plant nuclear genes among the 210 genes examined that descend from cyanobacterial genomes (white rectangles in Fig. 2). These genes became expressed by the eukaryotic transcription machinery and acquired transit peptides for reimport of the encoded products into the organelle of their genetic origin^{8,24}.

Why should plastids tend to donate genes to the nucleus at all? Nuclear primacy in gene regulation has many advantages²⁴, but population genetic factors may also be important. Notably, the effects of Muller's ratchet—the rapid accumulation of deleterious mutations in asexual populations—can be observed in metazoan mitochondrial genomes²⁵ and in endosymbiotic bacteria²⁶. Transfer of a gene from cpDNA to the nucleus should increase its rate of recombination and reduce its genetic load. This would generally favour nuclear fixation and could help to account for the degree of successful gene transfer indicated in Fig. 2. Chloroplast genomes furnish plant phylogenetics with a strong backbone and provide a simple model to study fragmented prokaryotic genomes dispersed across eukaryotic chromosomes^{8,27}. □

Methods

Complete sequences of *Zea mays* (X86563; *zea*), *Oryza sativa* (X15901; *oryz*), *Nicotiana tabacum* (S54304; *nic*), *Pinus thunbergii* (D17510; *pin*), *Marchantia polymorpha* (X04465; *mar*), *Euglena gracilis* (X70810; *eug*), *Porphyra purpurea* (U38804; *por*), *Odontella sinensis* (Z67753; *odo*), and *Cyanophora paradoxa* (U30821; *cya*) chloroplast genomes were retrieved from GenBank. Starting with the *Porphyra* sequence, BLAST searches were made with all designated encoded proteins and annotated open reading frames. Homologues from the other eight cpDNAs and from *Synechocystis* PCC6803 (ref. 7) (*syn*) were retrieved. For genes missing in at least one chloroplast genome by this search, local TFASTA searches were made against six-frame translations of the nucleotide sequences as a control. The set of 45 orthologous proteins encoded in all 10 genomes were aligned in individual files with PILEUP of the Genetics Computer Group (GCG) package, and written into PHYLIP²⁸ format with CLUSTALW²⁹.

The alignment of ten OTUs with 11,039 positions in each sequence was used as input for Fig. 1. For distance and parsimony analysis, the concatenated data was bootstrapped²⁸; for ML analysis, the REL bootstrap method¹⁰ was used. Using the 11,039-site concatenated data, we calculated the approximate lnL (ref. 10) for all 2,027,025 possible trees, and the best 1,000 trees by the approximate likelihood criterion were provided for full likelihood analysis. T1 had the highest likelihood. All trees that disrupted any of branches 3–7 were significantly ($\Delta\ln L > 5SE$) worse than T1. Therefore, branches 3–7 were defined, and the remaining 15 possible trees were more extensively examined using various subsets of the complete data (Table 1). Names of the 45 protein coding genes used for phylogenetic inference are given alphabetically, lengths of alignment used are in parentheses: *atpA* (494); *atpB* (491); *atpE* (237); *atpF* (182); *atpH* (81); *petB* (215); *petG* (38); *psaA* (754); *psaB* (738); *psaC* (82); *psaJ* (38); *psbA* (343); *psbB* (507); *psbC* (460); *psbD* (353); *psbE* (72); *psbF* (40); *psbH* (58); *psbI* (36); *psbJ* (42); *psbK* (46); *psbL* (39); *psbN* (43); *psbT* (31); *rpl2* (277); *rpl14* (124); *rpl16* (134); *rpl20* (115); *rpl22* (111); *rpl36* (38); *rpoB* (1098); *rpoC1* (689); *rpoC2* (1388); *rps2* (225); *rps3* (250); *rps4* (211); *rps7* (156); *rps8* (144); *rps11* (146); *rps12* (123); *rps14* (103); *rps18* (60); *rps19* (94); *ycf4* (175); *ycf9* (58).

In maize, it is estimated that about 25 sites in cpDNA are subject to RNA editing³⁰. Although the total degree of editing is not known for all OTUs considered here, it will only affect comparison of differentially edited codons at the small fraction of edited sites, and was therefore neglected. The concatenated alignment contains 4,180 constant and 6,859 polymorphic sites. Criteria for

indicating the presence of a functional nuclear gene (Fig. 2) were (1) expression (full-size, or nearly so, cDNA sequence characterized); (2) presence of a functional or putative transit peptide; and (3) greater similarity between cpDNA and nuclear homologues than between cyanobacterial and nuclear homologues. BLAST results apply to GenBank on 15 October 1997. The 32 amino-acid fragment of *ycf1* present in rice cpDNA (JQ0282) was neglected. To avoid confusion in Fig. 2, red and green⁸ type *rbcL* and *rbcS* are counted and mapped as two genes, rather than four. *PetL*, a 31-amino-acid long component of the cytochrome *b₆/f* complex, is too short to detect its *Synechocystis* homologue (gi1653694) among the highest BLAST scores, but was counted as being of cyanobacterial origin; the same applies for *ycf33* (67 amino acids long). BLAST search results, ML results for individual proteins, accession numbers for identified nuclear homologues, concatenate and individual alignments used in this study can be retrieved from 134.169.70.80/ftp/pub/incoming/.

Received 26 August 1997; accepted 4 February 1998.

1. Palenik, B. & Haselkorn, R. Multiple evolutionary origins of prochlorophytes, the chlorophyll *b*-containing prokaryotes. *Nature* **355**, 265–267 (1992).
2. Urbach, E., Robertson, D. L. & Chisom, S. W. Multiple evolutionary origins of prochlorophytes within the cyanobacterial radiation. *Nature* **355**, 267–270 (1992).
3. Helmchen, T. A., Bhattacharya, D. & Melkonian, M. Analyses of ribosomal RNA sequences from glaucocystophyte cyanobacteria provide new insights into the evolutionary relationships of plastids. *J. Mol. Evol.* **41**, 203–210 (1995).
4. Van de Peer, Y., Rensing, S., Maier, U.-G. & De Wachter, R. Substitution rate calibration of small subunit RNA identifies chlorarachniophyte endosymbionts as remnants of green algae. *Proc. Natl Acad. Sci. USA* **93**, 7744–7748 (1996).
5. Melkonian, M. Systematics and evolution of the algae: Endocytobiosis and the evolution of the major algal lineages. *Progr. Bot.* **57**, 281–311 (1996).
6. Halliwick, R. B. et al. Complete sequence of *Euglena gracilis* chloroplast DNA. *Nucleic Acids Res.* **21**, 3537–3544 (1993).
7. Kaneko, T. et al. Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res.* **3**, 109–136 (1996).
8. Martin, W. & Schnarrenberger, C. The evolution of the Calvin cycle from prokaryotic to eukaryotic chromosomes: a case study of functional redundancy in ancient pathways through endosymbiosis. *Curr. Genet.* **32**, 1–18 (1997).
9. Saitou, N. & Nei, M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425 (1987).
10. Adachi, J. & Hasegawa, M. Computer Science Monographs, No. 28. MOLPHY Version 2.3: Programs for Molecular Phylogenetics Based on Maximum Likelihood. (Institute of Statistical Mathematics, Tokyo, 1996).
11. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Contr.* **CA19**, 716–723 (1974).
12. Hedtke, B., Börner, T. & Weihe, A. Mitochondrial and chloroplast phage-like RNA polymerases in *Arabidopsis*. *Science* **277**, 809–811 (1997).
13. Naylor, G. J. P. & Brown, W. M. Structural biology and phylogenetic estimation. *Nature* **388**, 527–528 (1997).
14. Adachi, J. & Hasegawa, M. Model of amino acid substitution in proteins encoded by mitochondrial DNA. *J. Mol. Evol.* **42**, 459–468 (1996).
15. Lockhart, P. J., Steel, M. A., Hendy, M. D. & Penny, D. Recovering evolutionary trees under a more realistic model of sequence evolution. *Mol. Biol. Evol.* **11**, 605–612 (1994).
16. Lockhart, P. J., Larkum, A. W., Steel, M., Waddell, P. J. & Penny, D. Evolution of chlorophyll and bacteriochlorophyll: the problem of invariant sites in sequence analysis. *Proc. Natl Acad. Sci. USA* **93**, 1930–1934 (1996).
17. Gibbs, S. P. The chloroplast of *Euglena* may have evolved from symbiotic green algae. *Can. J. Bot.* **56**, 2883–2889 (1978).
18. Cavalier-Smith, T. Kingdom Protozoa and its 18 phyla. *Microbiol. Rev.* **57**, 953–994 (1993).
19. Henze, K. et al. A nuclear gene of eubacterial origin in *Euglena* reflects cryptic endosymbioses during protist evolution. *Proc. Natl Acad. Sci. USA* **92**, 9122–9126 (1995).
20. Cavalier-Smith, T. et al. Cryptomonad nuclear and nucleomorph 18S rRNA phylogeny. *Eur. J. Phycol.* **31**, 315–328 (1996).
21. Kowalik, K. V., Stöbe, B., Schaffran, I., Kroth-Pancic, P. & Freier, U. The chloroplast genome of a chlorophyll *a+c* containing alga, *Odontella sinensis*. *Plant Mol. Biol. Repr.* **13**, 336–342 (1995).
22. Löffelhardt, W. & Bohnert, H. J. in *The Molecular Biology of Cyanobacteria* (ed. Bryant, D. A.) 56–89 (Kluwer, Dordrecht, 1994).
23. Wakasugi, T. et al. Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of black pine *Pinus thunbergii*. *Proc. Natl Acad. Sci. USA* **91**, 9794–9798 (1994).
24. Herrmann, R. G. in *Eukaryotism and Symbiosis* (eds Schenck, H. E. A., Herrmann, R. G., Jeon, K. W. & Schwemmler, W.) 73–118 (Springer, Heidelberg, 1997).
25. Lynch, M. Mutation accumulation in transfer RNAs: Molecular evidence for Muller's ratchet in mitochondrial genomes. *Mol. Biol. Evol.* **13**, 209–220 (1996).
26. Moran, N. A. Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. *Proc. Natl Acad. Sci. USA* **93**, 2873–2878 (1996).
27. Martin, W. & Müller, M. The hydrogen hypothesis for the first eukaryote. *Nature* **392**, 37–41 (1998).
28. Felsenstein, J. PHYLIP (Phylogeny inference package) manual, version 3.5c (Univ. Washington, Seattle, Dept Genetics, 1993).
29. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTALW: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
30. Maier, R. M., Neckermann, K., Igloi, G. L. & Kössel, H. Complete sequence of maize chloroplast genome: Gene content, hotspots of divergence and the tuning of genetic information by transcript editing. *J. Mol. Biol.* **251**, 614–628 (1995).

Acknowledgements. We thank R. G. Herrmann for discussion, H. Phillippe for comments on the manuscript, the Rechenzentrum der Universität Braunschweig for use of computer facilities, and the Deutsche Forschungsgemeinschaft (W.M. and K.V.K.) and the Ministry of Education, Science, Sports and Culture of Japan (M.H.) for financial support. B.S. is the recipient of a postdoctoral stipend from the DFG; V.G. is the recipient of a Ph.D. stipend from DAAD.

Correspondence and requests for materials should be addressed to W.M. (e-mail: w.martin@tu-bs.de).

Regulation of leaf initiation by the *terminal ear 1* gene of maize

Bruce Veit^{*†‡}, Steven P. Briggs[§], Robert J. Schmidt[‡], Martin F. Yanofsky[‡] & Sarah Hake[†]

^{*} Institute of Molecular Biosciences, Massey University, Palmerston North, New Zealand

[†] Plant Gene Expression Center, USDA/UC Berkeley, Albany, California 94710, USA

[‡] Department of Biology, University of California, San Diego, California 92093, USA

[§] Pioneer Hi-Bred, International Inc., Johnston, Iowa 50131, USA

Higher plants elaborate much of their architecture post-embryonically through development initiated at the tips of shoots^{1,2}. During vegetative growth, leaf primordia arise at predictable sites to give characteristic leaf arrangements, or phyllotaxies^{3,4}. How these sites are determined is a long-standing question^{5,6} that bears on the nature of pattern-formation mechanisms in plants. Fate-mapping studies in several species indicate that each leaf primordium becomes organized from a group of 100–200 cells on the flank of the shoot apex⁷. Although molecular studies indicate that the regulated expression of specific homeobox genes plays some part in this determination process^{8–11}, mechanisms that regulate the timing and position of leaf initiation are less well understood. Here we describe a gene from maize, *terminal ear 1*. Patterns of expression of this gene in the shoot and phenotypes of mutants indicate a role for *terminal ear 1* in regulating leaf initiation. The

te1 gene product contains conserved RNA-binding motifs, indicating that it may function through an RNA-binding activity.

The *terminal ear 1* (*te1*) mutant¹² is named for the earlike inflorescence that forms in place of the normal terminal tassel¹³ (Fig. 1a). The unusual appearance of this mutant derives largely from the presence of abnormally short internodes (segments of the stem that separate adjacent leaves) that cause vegetative leaves to overlap and enclose the tassel, much like the husk leaves that normally enclose the axillary ear (Fig. 1b, c). In some genetic backgrounds, varying degrees of tassel feminization further contribute to an earlike appearance. The initiation and development of the leaves themselves is also affected, with irregular deviations from the normal alternate phyllotaxy (successive leaves positioned 180° apart on the shoot axis) and characteristically narrow leaves in which the blade has a more perpendicular orientation to the shoot axis than is normal (Fig. 1d). Some leaves also have large-scale pattern defects, in which the normal single midrib may be absent, replaced by two midribs, or asymmetrically positioned with respect to the blade margins (Fig. 1d).

Comparisons of *te1* mutant and normal siblings reveal differences in stem architecture. The mutant plants exhibit shorter internodes over the entire length of the stem, leading to decreased plant stature (Table 1). Short internodes are most apparent in the upper half of the plant, although there is also an increase in the number of short internodes normally seen at the base of the plant (Table 1). In most individuals, one or more abnormally short internodes may be interspersed between internodes of relatively normal length, particularly in the upper third of the plant (Fig. 1b, c, Table 1). The decreased internode-length phenotype is accompanied by an increase in the total number of internodes formed (Table 1). This increase can be attributed to an increased rate of leaf initiation, especially during the latter stages of development when abnormally short internodes and aberrant leaves are most likely to form (Table 1).

To investigate the mechanism by which the wild-type *te1* gene affects shoot development, we cloned the gene by transposon tagging with *Robertson's Mutator*¹⁴ (see Methods). The longest complementary DNA associated with the tagged interval, of 2,425 base pairs (bp) in length, includes an open-reading frame encoding 656 amino acids (Fig. 2a). This protein shows significant similarity to Mei2 (ref. 15) (Fig. 2b, c), an RNA-binding protein from

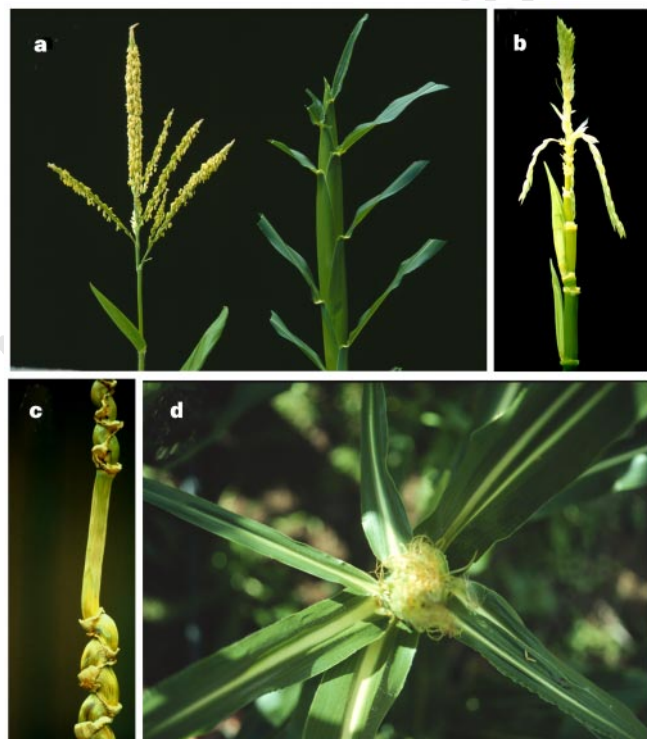


Figure 1 Explanation of the *te1* mutant phenotype. **a**, Comparison of a normal (left) and *te1-1* mutant (right) tassel, showing the overlapping leaves that enclose the mutant tassel. **b**, Upper part of a mature *te1-1* mutant plant, with leaves removed to illustrate the irregular pattern of short vegetative internodes and a relatively normal tassel. **c**, Stalk from a *te1* mutant plant, showing a normal length internode interrupting a series of abnormally short internodes. **d**, Top view of a *te1-1* mutant plant, showing deviations from the normal alternate phyllotaxy and leaves with double midribs.

Table 1 Phenotype analysis of *terminal ear 1* mutants

Character	Normal	<i>Terminal ear 1</i>
Plant height (cm)*	189 ± 6.4	105 ± 9.1
Internode length*		
Whole plant	9.2 ± 5.2	3.8 ± 4.1
Top third	10.8 ± 2.1	3.6 ± 4.8
Middle third	14.1 ± 2.2	7.9 ± 3.6
Basal third	6.2 ± 5.6	2.7 ± 2.5
Length dissimilarity*		
Whole plant	0.11	0.27
Top third	0.03	0.48
Middle third	0.04	0.28
Basal third	0.18	0.14
Internode no.*		
Whole plant	20.4 ± 1.1	27.4 ± 0.9
Top third	5.8 ± 1.0	9.7 ± 1.3
Middle third	4.6 ± 0.1	4.4 ± 0.56
Basal third	10.0 ± 0.3	13.4 ± 0.5
No. of leaves initiated†		
10 days	8.7 ± 0.5	10.2 ± 0.4
17 days	11.0 ± 0.7	14.2 ± 1.3
24 days	14.6 ± 1.1	20.2 ± 1.3

*Averaged data from comparisons of mature *te1* mutant and normal siblings (five each) as described in Methods. Length dissimilarity (l.d.) provides a measure of the degree to which adjacent internodes differ in length, with l.d. = $|l^x - l^{x+1}| / (l^x + l^{x+1})$, where l^x equals the length of a selected internode and l^{x+1} equals the length of the adjacent internode above. L.d. may range from 0 (identical lengths) to a maximum limit of 1 (very dissimilar lengths). † Average leaf numbers present at different ages (see Methods).

Schizosaccharomyces pombe that is required for both premeiotic DNA synthesis and the first reductional division of meiosis. The similarity is highest in the three regions in both protein sequences that encode RNA-recognition motifs (RRMs)^{16,17}, with 24%, 31% and 40% identity found for RRM1, 2 and 3, respectively (Fig. 2c). The *Mu8* transposon elements associated with the two alleles *tel-mum1* and *tel-mum2* are inserted at sites 256 bp apart in the third exon. A third allele, *tel-mum3*, is associated with a *Mutator* insertion into a small, 102-bp intron that separates exons 1 and 2.

In initial studies of gene expression, using northern blot analysis of poly(A)⁺ RNA, we detected a 2.8-kilobase (kb) transcript in shoot apices from wild-type (WT) plants, but not in those from plants homozygous for the *tel-1* mutant allele. No transcripts were detected in root tips, immature ears (husk leaf primordia were not included), or expanded seedling leaves of WT plants (B.V., unpublished observations). More detailed expression studies by *in situ* hybridization showed that *tel* transcripts accumulated in the WT vegetative shoot apex in a series of semicircular bands, 4–5 cells wide, that extend circumferentially around the dome-shaped meristem (Fig. 3a). In median longitudinal section (Fig. 3b), these bands appear as a series of patches in epidermal and subepidermal layers that begin 10–20 cells below the shoot tip and alternate from side to side down the flanks of the shoot apex. This alternating pattern has a consistent relationship to leaf primordia, which begin their overt development in the gaps between the arms of the semicircular expression bands where *tel* expression is lowest (Fig. 3c). This same basic pattern is seen in 17-day-old embryos (Fig. 3d). There was no signal above background in shoot apices of *tel-1* homozygotes (Fig. 3e). These mutant shoot apices also exhibited an abnormal geometry, producing leaf initials higher on the apical dome than is normal. Decreased levels of *KNOTTED1* transcripts, an early marker of leaf founder cell identity^{8,11} in normal plants, show a close association with the abnormally positioned primordia (Fig. 3f).

Our analyses indicate that the *tel* gene is important in the early stages of leaf development. An increased frequency of leaf initiation associated with loss of *tel* function indicates that the gene may normally limit this process. The irregular phyllotaxy of the mutant, the abnormal geometry of its shoot apex and the associated changes in the expression of *KNOTTED1* suggest that the *tel* gene acts early in the initiation process, when the angular and longitudinal positioning of leaf is being determined. These phenotypes and the normal positioning of initiation sites in regions where *tel* levels are lowest suggest a model in which *tel* inhibits cells in which it is expressed from acting as organizers of leaf development. Loss of *tel*

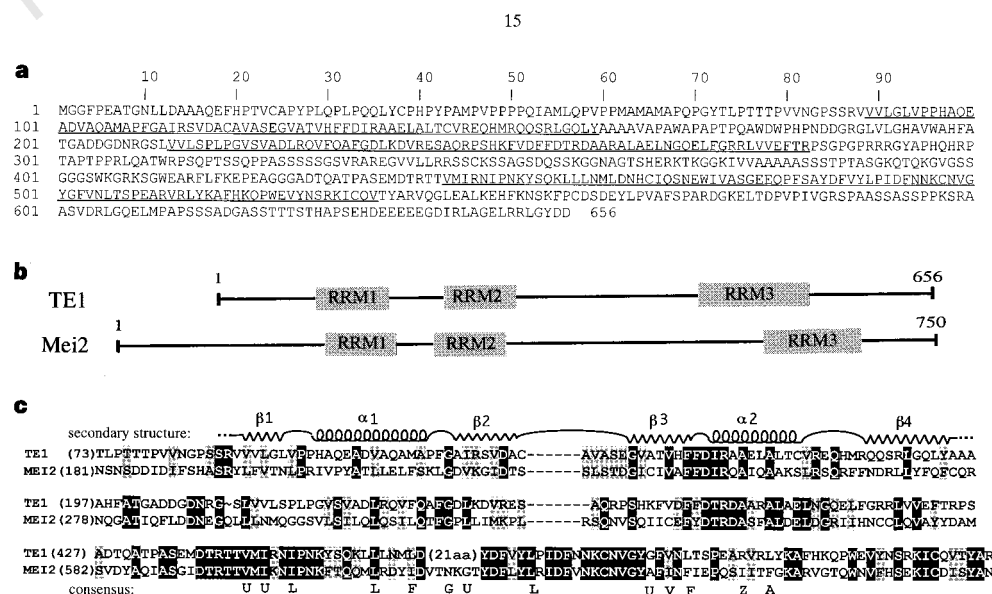
function would allow cells to act as organizers in normally restricted regions of the apex, leading to precocious and/or abnormally positioned initiation events. Observed pattern defects involving the midrib of the leaf may be another outcome of these less constrained initiation events. Results of molecular and genetic studies are consistent with cells in the midvein domain acting as organizers in the leaf-initiation process¹⁰.

The *tel* gene also seems to influence later phases of development, as indicated by altered leaf and internode development of the mutant. These changes may be an indirect result of precocious or abnormally positioned leaf-initiation events that would lead to some elements of the shoot developing from abnormally small numbers of founder cells. It may be significant that the aberrant, short internodes typically show a pronounced length asymmetry, with the short side located directly below the site of initiation (the midrib-containing region) of the leaf above. Such an asymmetry might result if cells normally partitioned to the incipient internode were instead diverted into the adjacent, clonally related leaf⁸. Alternatively, the altered development of the mutant leaf and internode might reflect an ongoing requirement for the *tel* gene product to regulate and promote cell division within these structures. This interpretation is consistent with mutant internode development being attenuated most severely on the side where *tel* levels would normally be highest.

The earlike appearance of the terminal inflorescence of *tel* plants prompts us to consider models in which *tel* would act to switch between the ear and tassel programmes of development. Such a function has been proposed for teosinte branched 1, a recently cloned gene¹⁹ whose expression seems to promote the conversion of elongated tassel-tipped branches, typical of the ancestral teosinte, into the compressed feminized shoot that constitutes the ear. Although the earlike appearance of the *tel* tassel might indicate that the gene normally promotes a tassel programme of development over a default ear programme, the patterned expression of the gene within the normal vegetative shoot apex and the leaf-initiation and pattern defects of the mutant favour a direct role for *tel* in controlling the leaf-initiation process itself. The expression of *tel* in WT axillary shoot apices that are initiating husk leaves (B.V., unpublished observations) is further evidence that *tel* has a more direct influence on leaf initiation, rather than globally determining shoot identity. Our model, however, does not simply explain the earlike feminization of the *tel* tassel seen in some genetic backgrounds. This feminization could reflect a distinct physiology associated with the short internodes that form below the tassel of the *tel* plant. A similar explanation has been offered for the

Figure 2 Predicted TE1 protein structure. **a**,

Amino-acid sequence resulting from translation of the 2,425-bp *tel* cDNA. **b**, Comparison of the structures of the TE1 and Mei2 proteins, showing the location of conserved RRM1, 2 and 3. **c**, Alignments of three RRM1, 2 and 3 found in both the TE1 and the Mei2 proteins. Identical amino acids are shown on a black background, and conserved pairs are shown on a grey background. All three pairs are aligned with a schematic diagram of a canonical RRM secondary structure, shown at the top. Consensus residues deduced from comparisons of other RRM-containing genes^{16,21} are shown on bottom line (U, uncharged residues; Z, U + S + T¹⁶).



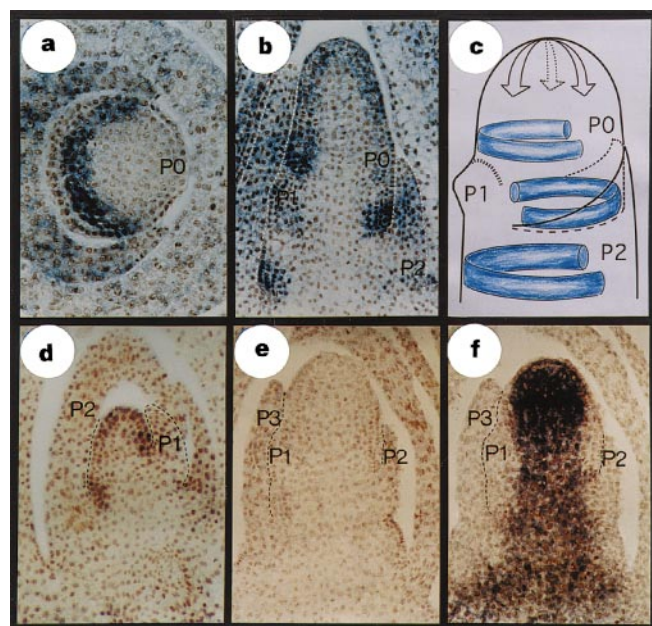


Figure 3 Analysis of *te1* expression. The figure shows accumulation patterns of either *te1* (a-e) or *KNOTTED1* (f) transcripts in shoot apices. **a**, Transverse cross-section of WT shoot apex 0.2 mm below the shoot tip. **b**, Median longitudinal cross-section of WT shoot apex. **c**, Schematic diagram of the patterns shown in **a**, **b**. **d**, Median longitudinal section of a 17-day-old embryo. **e**, **f**, Near-adjacent longitudinal sections hybridized with probes specific for either *te1* (**e**) or *KNOTTED1* (**f**). All panels except **d** represent shoot apices at three weeks after germination. Successively older leaf primordia are labelled P0-P3, with labels positioned over where primordia would first emerge.

feminized pattern of development of the short-ear-bearing axillary shoots of normal maize²⁰.

The similarity of the TE1 protein to Mei2 provides some clues to the mechanism by which it acts. Members of the larger class of RNA-binding proteins to which Mei2 belongs exhibit several interactions with RNAs, including regulation of differential splicing, localization and stabilization of messenger RNAs²¹. *FCA*, a gene that controls flowering in *Arabidopsis*²², encodes one such RNA-binding protein. *FCA* transcripts show a pattern of differential splicing that is indicative of a post-transcriptional control mechanism. It is unknown whether Mei2, to which TE1 is more closely related, acts in a similar way; however, genetic and molecular analyses indicate that Mei2 binds to a non-coding poly(A)⁺ RNA²³ to enable progression through the first meiotic division and interacts with an as yet undefined RNA to enable premeiotic DNA synthesis. The function of the TE1 protein is unrelated to meiosis, but its similarity to Mei2 may reflect conserved mechanisms defined at a more basic level. As changes in the polarity and frequency of cell division are associated with the earliest stages of leaf initiation, perhaps the TE1 protein functions as part of a regulatory mechanism to control these processes. □

Methods

Phenotypic analysis. Sibling plants for both mature-plant and leaf-initiation studies were from a line in which the *te1-1* reference allele had been introgressed into a standard B73 inbred genetic background by three generations of backcrossing to decrease variability from unlinked modifiers. Plants used in the study were progeny of self-fertilized heterozygotes. Leaf numbers were determined by microscopic dissection of five or more plants of each class at 1-week intervals, with all leaves and leaf primordia counted towards the total number of leaves initiated.

Molecular cloning. Two independent alleles, *te1-mum1* and *te1-mum2*, were identified in an untargeted transposon mutagenesis experiment. Both alleles

were found to co-segregate with a 3-kilobase (kb) *Bcl1* restriction fragment that contained the 1.4-kb *Mu8* transposon¹³, with no recombinations detected in >100 chromosomes. A non-repetitive 0.8-kb *Hinf1* DNA fragment flanking the *Mu8* element of *te1-mum1* was used to isolate a 12-kb *Asp718* restriction fragment from a genomic library from the WT B73 inbred. This larger fragment, which spans the *Mu8* insertion sites of both *te1-mum1* and *te1-mum2*, was used as a hybridization probe to identify ten homologous cDNAs from 10⁶ primary recombinants screened in a ZAPII library (Stratagene) made with poly(A)⁺ RNA from normal vegetative maize shoot apices. The longest of these, a 2,425-bp cDNA, was sequenced and found to include an open-reading frame (ORF) encoding 656 amino acids (Fig. 3a). The presence of an in-frame stop codon located 5' to the initiating ATG and nearby stop codons in the other two reading frames suggests that the long ORF of this cDNA encodes the functional, full-length protein. To demonstrate that this region corresponds to the *te1* locus, we analysed progenitor material. Parental material was not available for the self-pollinated plants in whose progeny the new *te1-mum* alleles were first recognized in a homozygous state. Therefore, we analysed the progeny of self-pollinated siblings of these plants to infer the chromosomal constitution of the parents. Southern blot analysis of *te1-mum1* in pools of six progeny from each of six siblings showed no evidence of the *Mu8* element, indicating that insertion of *Mu8* probably produced the new allele. We analysed *te1-mum3* similarly by PCR; the coincidence of the *Mutator* element insertion and the appearance of the new allele further demonstrates that this disrupted gene corresponds to the *te1* locus.

In situ hybridization. Paraffin-sectioned material was hybridized with digoxigenin-labelled probes corresponding to the antisense strand of the 2.4-kb cDNA. The fixation of tissues, hybridizations, washes and detection steps were performed as described¹¹ except that hybridizations were performed at 60–65 °C to compensate for the high GC content of the *te1* gene.

Received 8 December 1997; accepted 3 March 1998.

- Walbot, V. On the life strategies of plants and animals. *Trends Genet.* **1**, 165–169 (1985).
- Steeves, T. A. & Sussex, I. M. *Patterns in Plant Development* (Cambridge Univ. Press, Cambridge, 1989).
- Sussex, I. M. Developmental programming of the shoot meristem. *Cell* **56**, 225–229 (1989).
- Smith, L. G. & Hake, S. The initiation and determination of leaves. *Plant Cell* **4**, 1017–1027 (1992).
- Church, A. H. *On the Relation of Phyllotaxy to Mechanical Laws* (Williams and Norgate, London, 1904).
- Richards, F. J. Phyllotaxis: its quantitative expression and relation to growth in the apex. *Phil. Trans. R. Soc. Lond. B* **236**, 509–564 (1951).
- Poethig, R. S. Clonal analysis of cell lineage patterns in plant development. *Am. J. Bot.* **74**, 581–594 (1987).
- Smith, L., Greene, B., Veit, B. & Hake, S. A dominant mutation in the maize homeobox gene, *Knotted-1*, causes its ectopic expression in leaf cells with altered fates. *Development* **116**, 21–30 (1992).
- Long, J. A., Moan, E. L., Medford, J. I. & Barton, M. K. A member of the KNOTTED class of homeodomain proteins encoded by the *STM* gene of *Arabidopsis*. *Nature* **379**, 66–69 (1996).
- Scanlon, M. J., Schneeburger, R. G. & Freeling, M. The maize mutant *narrow sheath* fails to establish leaf margin identity in a meristematic domain. *Development* **122**, 1683–1691 (1996).
- Jackson, D., Veit, B. & Hake, S. Expression of maize *KNOTTED1* related homeobox genes in the shoot apical meristem predicts patterns of morphogenesis in the vegetative shoot. *Development* **120**, 405–413 (1994).
- Matthews, D. L., Grogan, C. O. & Manchester, C. E. Terminal ear mutant of maize (*Zea mays*). *J. Agric. Sci.* **82**, 433–435 (1974).
- Kieselbach, T. A. *The Structure and Reproduction of Corn* (Univ. Nebraska, Lincoln, 1949).
- Chandler, B. L. & Hardeman, K. J. The *Mu* elements of *Zea mays*. *Adv. Genet.* **30**, 77–122 (1992).
- Watanabe, Y., Iino, Y., Furuhata, K., Shimoda, C. & Yamamoto, M. The *S. pombe* *mei2* gene encoding a crucial molecule for commitment to meiosis is under the regulation of cAMP. *EMBO J.* **7**, 761–767 (1988).
- Birney, E., Kumar, S. & Krainer, A. R. Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. *Nucleic Acids Res.* **21**, 5803–5816 (1993).
- Watanabe, Y., Shinozaki-Yabana, S., Chikashige, Y., Hiraoka, Y. & Yamamoto, M. Phosphorylation of RNA-binding protein controls cell cycle switch from mitotic to meiotic in fission yeast. *Nature* **386**, 187–190 (1997).
- McDaniel, C. & Poethig, R. S. Cell-lineage patterns in the shoot apical meristem of the germinating maize embryo. *Planta* **175**, 13–22 (1988).
- Doebley, J., Stec, A. & Hubbard, L. The evolution of apical dominance in maize. *Nature* **386**, 485–488 (1997).
- Itis, H. From *teosinte* to maize: the catastrophic sexual transmutation. *Science* **222**, 886–894 (1983).
- Burd, C. G. & Dreyfuss, G. Conserved structures and diversity of functions of RNA-binding proteins. *Science* **265**, 615–621 (1994).
- Macknight, R. et al. *FCA*, a gene controlling flowering time in *Arabidopsis*, encodes a protein containing RNA-binding domains. *Cell* **89**, 737–745 (1997).
- Watanabe, Y. & Yamamoto, M. *S. pombe* *mei2+* encodes an RNA-binding protein required for premeiotic DNA synthesis and meiosis I, which cooperates with a novel RNA species *meiRNA*. *Cell* **78**, 487–498 (1994).

Acknowledgements. This work was supported by NSF and USDA grants (to S.H.), a USDA grant (to R.J.S. and M.F.Y.) and grants from Pioneer Hi-Bred Int. and the Marsden Foundation (to B.V.). Technical assistance in cloning the *te1* gene was provided by E. Robbie. *te1-mum2* was a gift from P. Chomet. GenBank accession number: AF047852.

Correspondence and requests for materials should be addressed to B.V.

Neonatal lesions of the medial temporal lobe disrupt prefrontal cortical regulation of striatal dopamine

Richard C. Saunders^{*†}, Bhaskar S. Kolachana^{*}, Jocelyne Bachevalier[‡] & Daniel R. Weinberger^{*}

^{*} Clinical Brain Disorders Branch, NIMH Neuroscience Center at St Elizabeth's Hospital, Washington DC 20032, USA

[‡] Department of Neurobiology and Anatomy, University of Texas, Houston, Texas 77030, USA

[†] Present address: Clinical Brain Disorders Branch, NIMH, Building 49, Room 1B80, 49 Convent Drive, Bethesda, Maryland 20892-4415, USA

The effects of early brain damage are often, but not always, milder than the effects of comparable damage in adults, depending on the age at which injury occurred, the region of the brain damaged, and the brain functions involved^{1–7}. Studies of the impact of early brain damage have generally focused on functions primarily associated with the neural structures injured, even though the development and function of distant but interconnected neural systems might also show effects. Here we examine the regulation of striatal dopamine by the dorsolateral prefrontal cortex, in adult monkeys that had had either neonatal or adult lesions of the medial–temporal lobe and in normal animals. We use microdialysis to measure the dopamine response in the caudate nucleus after the infusion of amphetamine into the dorsolateral prefrontal cortex. Normal animals and those with adult lesions showed a reduction in dopamine overflow; in contrast, monkeys with neonatal lesions showed increased dopamine release. Thus, early injury to the primate medial–temporal lobe disrupts the normal regulation of striatal dopamine activity by the dorsolateral prefrontal cortex during adulthood. Early focal lesions may

have substantial and long-lasting impacts on the function of a distant neural system.

In non-human primates, studies of prenatal and neonatal frontal lobe lesions have been focused on the development of frontal-related behaviours such as working memory^{1,3}, whereas studies of neonatal medial–temporal limbic lesions, including lesions of the hippocampus, have been focused on memory or emotional behaviours that tend to be affected by similar lesions made in adults^{4,5}. However, changes in neural connectivity associated with injury to the developing brain could have widespread functional implications. This is indicated by experiments in the rat that show that neonatal hippocampal injury results in behavioural hyper-responsivity of striatolimbic dopaminergic systems to environmental and pharmacological stresses⁶. Also, some monkeys with early limbic lesions show increased locomotor stereotypies not seen in animals with similar adult lesions⁷. We questioned whether a developmental lesion of temporal lobe limbic structures (Fig. 1) could affect adult mechanisms of dopaminergic regulation subserved by the prefrontal cortex in the non-human primate, which has a diverse and developed prefrontal cortex composed of neocortical areas not present in lower mammals⁸. Glutamatergic efferents of the dorsolateral prefrontal cortex (DLPFC) have been shown to regulate caudate dopamine release in the monkey^{9,10}.

Basal dopamine levels in the caudate nucleus did not differ significantly between the three groups of monkeys studied (normal animals: 137.2 ± 22.7 fmol per $15 \mu\text{l}$, $N = 17$ probes; adult lesion: 106.5 ± 18.4 fmol per $15 \mu\text{l}$, $N = 12$; neonatal lesion: 121.7 ± 18.9 fmol per $15 \mu\text{l}$; $N = 12$; $F = 1.40$, degree of freedom (d.f.) = 2, $P = 0.26$). Amphetamine infusion into the DLPFC resulted in large changes in dopamine overflow in the caudate nucleus (Fig. 2). Analysis of variance (ANOVA) revealed no overall group ($F = 1.8$, d.f. = 2, $P > 0.3$) or probe ($F = 0.88$, d.f. = 2, $P > 0.5$) effects. However, there was an effect of time ($F = 5.7$, d.f. = 8, $P < 0.005$) and a strong group–time interaction ($F = 9.6$, d.f. = 16, $P < 0.0001$). This statistical interaction effect emerged because the lesion of the medial–temporal lobe (MTL) had markedly different physiological effects on the regulation of dopamine release in the caudate nucleus by the prefrontal cortex. Normal

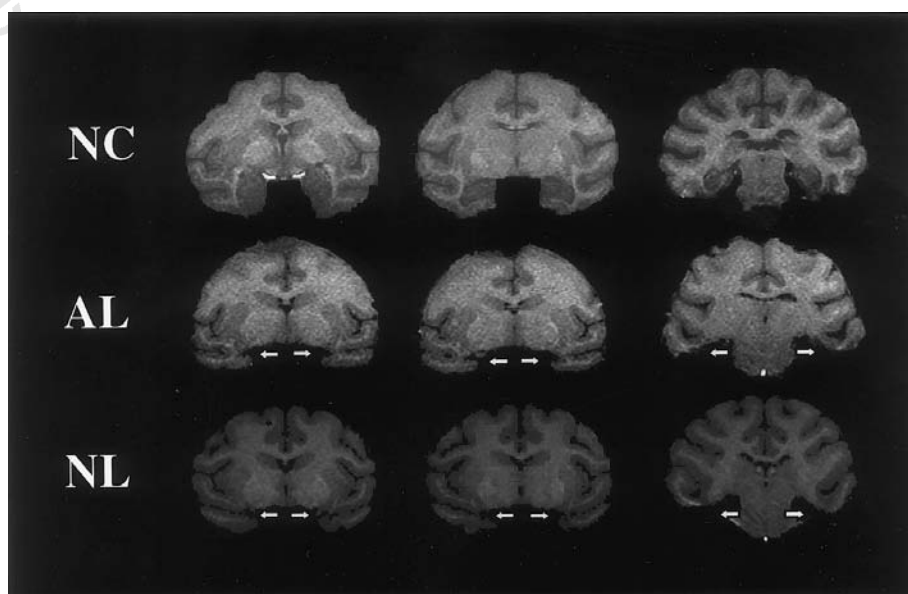


Figure 1 Magnetic resonance images of monkeys, showing rostral-to-caudal coronal sections. The first row shows images from a normal monkey (NC, normal control) with the amygdala, rhinal cortical areas and hippocampal formation (left to right) intact. The second row illustrates a monkey in which the MTL was removed in adulthood (adult lesion, AL); in the third row are shown images from a

monkey with a neonatal lesion (NL). The white arrows point to the sites of the removal. In all monkeys with lesions there was some slight bilateral damage to the posterior part of the inferior temporal cortex, lateral to the intended lesion. There were no obvious differences in the lesions between the two groups.

monkeys and those lesioned as adults showed a decrease in caudate dopamine overflow after D-amphetamine infusion into the DLPFC (Fig. 2), consistent with earlier results¹⁰. In contrast, the monkeys lesioned as neonates showed the opposite effect, an increase in dopamine release (Fig. 2). The neonatal limbic lesion thus seems to result in a reversal of the direction of prefrontal cortical regulation of caudate dopamine levels.

This aberrant regulatory response may result from many factors, including the different effects of D-amphetamine on catecholamines in the DLPFC, on release of glutamate and/or γ -aminobutyric acid (GABA) in the DLPFC, on glutamatergic activity of prefrontal-caudate projections, and/or on the regulation of dopamine cell firing patterns at the level of the midbrain.

Infusion of D-amphetamine resulted in similar increases in dopamine levels in the DLPFC in all three groups of animals (normal animals: 7.4 ± 1.1 fmol per $15 \mu\text{l}$, $N = 8$ probes; adult

lesion: 7.8 ± 2.2 fmol per $15 \mu\text{l}$, $N = 8$; neonatal lesion: 6.8 ± 0.7 fmol per $15 \mu\text{l}$, $N = 8$) (Fig. 3). The magnitude of these increases was similar to that reported after systemic administration of D-amphetamine¹¹. Glutamate overflow in the prefrontal cortex did not differ significantly between the groups either (there was no main group effect, $F = 1.39$, d.f. = 2, $P = 0.27$, or a time-group interaction, $F = 0.48$, d.f. = 16, $P = 0.95$), and nor did GABA levels ($F = 0.22$, d.f. = 2, $P = 0.81$). Furthermore, glutamate overflow in the caudate nucleus did not differ between the groups and tended to be relatively stable (Fig. 4) (normal animals: 10.1 ± 2.5 ng per $10 \mu\text{l}$, $N = 8$ probes; adult lesion: 14.2 ± 4.4 ng per $10 \mu\text{l}$, $N = 8$; neonatal lesion: 12.0 ± 2.2 ng per $10 \mu\text{l}$, $N = 8$; group effect $F = 0.18$, d.f. = 2, $P = 0.84$; time-group interaction $F = 1.79$, d.f. = 16, $P = 0.36$). It is, however, possible that some small change in one group might have been masked by the overall declining baseline levels, although because of the shape of the curves and the magnitude of change, this seems doubtful. Although we did not examine activity of glutamatergic projections to the midbrain, they would probably not be functionally differentiated from cortical-striatal projections as they seem to originate from a similar group of neurons¹². Thus, the lesion effect cannot be explained by a simple differential pharmacological action of amphetamine at the level of dopamine-releasing terminals, or by different glutamatergic or GABAergic neuronal responses of the prefrontal cortex.

Although we cannot specify the precise circuitry by which the early limbic damage affects the caudate dopamine response to prefrontal challenge, the data indicate a polysynaptic mechanism involving prefrontal connections with dopamine cell bodies within the midbrain, that is, the ventral tegmental area and/or the substantia nigra¹³. The effects of prefrontal cortical manipulation on striatal dopamine levels have been shown to be blocked by administration of glutamate antagonists into the midbrain but not into the striatum^{13,14}. These regulatory connections might be affected by the impact of the early lesion at various sites^{15,16}. In normal animals there are projections from these medial temporal limbic areas to the DLPFC¹⁷ and, to a lesser extent, to the caudate nucleus and midbrain^{18–20}. Our results indicate that there may not be a single pharmacological disruption of these projections, but the possibility of synaptic modification at the level of the DLPFC needs to be explored. We have shown, in these animals²¹, that there is a neurochemical change (that is, a reduction in *N*-acetyl-aspartate nuclear-magnetic-resonance signals) in the

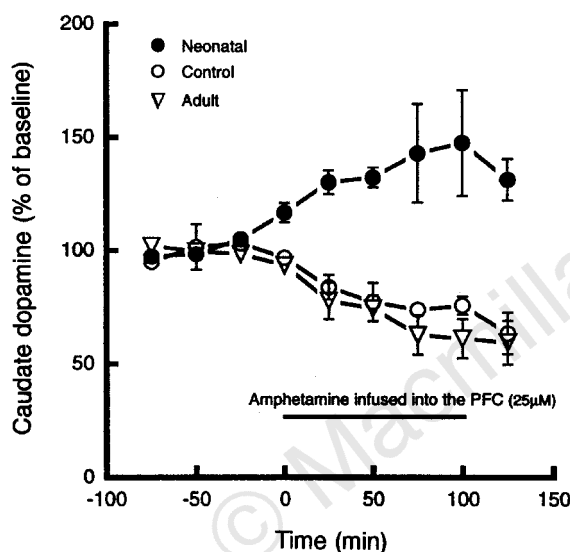


Figure 2 Monkeys with adult lesions and those with neonatal limbic lesions showed opposite effects on caudate dopamine levels associated with D-amphetamine infusion into the prefrontal cortex (PFC).

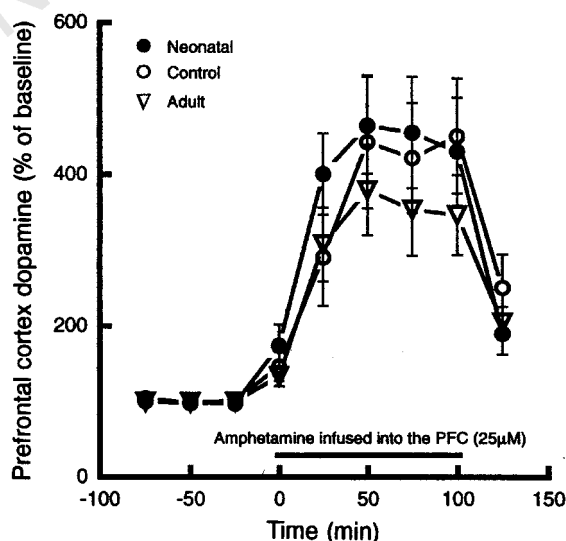


Figure 3 The dopamine response in the prefrontal cortex (PFC) to amphetamine infusion into the PFC did not differ significantly between the three groups.

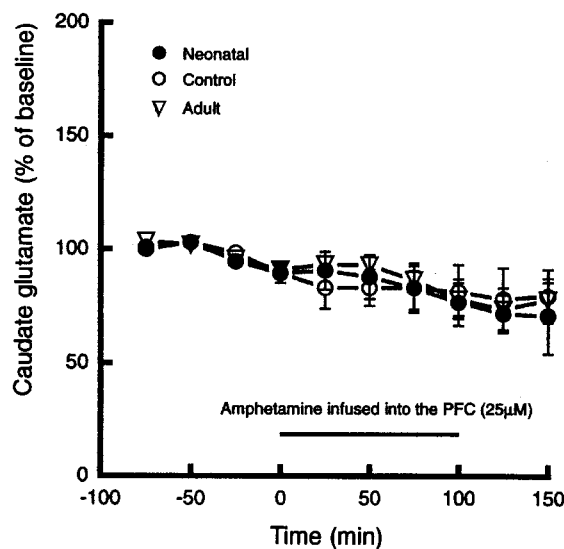


Figure 4 Glutamate levels in the caudate nucleus in response to prefrontal cortical (PFC) amphetamine infusion.

DLPFC associated with the neonatal MTL lesion but not with the adult lesions.

Our results indicate that some cases of early focal injury result in physiological alterations not predicted by effects of adult lesions or by knowledge of functions traditionally linked to the lesion site. Although behavioural data indicate that these animals have impaired recognition memory⁴, as is normally associated with lesions of this region, other cognitive changes and biological effects such as neuroanatomical reorganization and receptor variations have not been examined. Our results may have implications for understanding how certain environmental stimuli (such as stress) may have unexpected or problematic effects on brain function in the context of developmental injury. Indeed, monkeys with the neonatal lesions of the MTL show increased locomotor stereotypies, especially in stressful situations such as social interactions⁷. A clinical disorder in which such adverse stress effects are often seen is schizophrenia, a condition associated with several phenomena analogous to those seen in our MTL monkeys, including possible developmental injury to the limbic temporal lobe²², abnormal function of the DLPFC²³, putative dysregulation of dopaminergic systems^{24,25}, stress-related clinical decomposition²⁶, and exaggerated striatal dopaminergic responses to amphetamine^{27,28}.

Methods

Subjects and surgery. There were three animal groups, each consisting of two adult monkeys (*Macaca mulatta*) of 9 or 10 years. The MTL removal included the amygdala, entorhinal cortex, hippocampal formation, and parahippocampal gyrus. Neonatal MTL removals were made during the first three weeks of life. In the adult animals there was a minimum of two years from the date of surgery to the beginning of the microdialysis experiments. The extent of the lesion was assessed using magnetic resonance imaging (Fig. 1). Normal animals and those with neonatal lesions were reared in almost identical social and physical environments. Each monkey was a subject in two or three separate dialysis sessions.

In vivo microdialysis. For a typical dialysis session (defined as a complete microdialysis experiment performed generally for a period of 8–10 h), three monkeys, one from each group, were initially sedated with ketamine (intramuscular ketamine hydrochloride, 10 mg kg⁻¹), intubated, lightly anaesthetized with gas isoflurane (1–2%) and placed in a head holder. Vital signals were monitored throughout the experimental session. Microdialysis probes were constructed and secured into the targeted areas as previously described¹⁰. Three to four probes were positioned into the caudate nucleus in one hemisphere and three to four probes into the medial bank of the principal sulcus, ipsilaterally. Artificial cerebrospinal fluid (Na⁺ 147 mM, K⁺ 3 mM, Ca²⁺ 1.3 mM, Mg²⁺ 1.0 mM, Cl⁻ 155 mM, buffered with 1.0 mM phosphate, pH 7.3–7.4) supplemented with 0.15 mM ascorbate was continuously perfused through the probes at 1.0 µl min⁻¹ flow rate. Dialysis samples (25 µl) were collected every 25 min, split into two fractions of 10 µl and 15 µl and immediately frozen on dry ice. They were assayed for dopamine, glutamate, and GABA using microbore HPLC coupled to electrochemical detection^{10,29}. As these animals were anaesthetized throughout the procedure we cannot rule out an effect of anaesthesia on absolute transmitter levels. However, we know of no reason why anaesthesia would differentially affect the direction of the changes observed in the three groups of monkeys.

Basal levels of caudate dopamine overflow were established 4 h after probe insertions. Infusion of D-amphetamine into each of the DLPFC probes served as a pharmacological model of the increased release of synaptic concentrations of biogenic amines seen in the DLPFC during certain experiential stresses. D-amphetamine (25 µM) dissolved in artificial cerebrospinal fluid was infused for 100 min and was followed with standard artificial cerebrospinal fluid for an extra 4 h. Dialysis probes in the caudate nucleus were without interruption perfused with standard artificial cerebrospinal fluid throughout. Only samples with stable baseline levels were included in subsequent analyses.

Analysis. As a result of the discrete topography of cortical–striatal projections, we expected occasional anatomical mismatch between cortical and caudate probes and occasional negative results on this basis. Therefore, we used an initial statistical analysis to determine whether the frequency of the changes

seen after prefrontal cortical D-amphetamine administration might have occurred by chance. The differences in the frequency with which a manipulation in the DLPFC resulted in a consistent change in dopamine levels compared with the frequency of no change or a change opposite in direction to the predominant change (for example, an increase in the case of normal monkeys) were significant in all three groups (Fisher exact, $P < 0.01$ in each case). Further comparison of the absolute data (fmol µl⁻¹) from probes showing a change was carried out by a three-way ANOVA, with lesion status as the between-group variable, and time and probe (equalized in number across groups) as repeated measures. Bonferroni correction was used for multiple comparisons.

Received 27 October 1997; accepted 3 March 1998.

- Goldman-Rakic, P. S., Isseroff, A., Schwartz, M. L. & Bugbee, N. M. in *Handbook of Cognitive Development: Biology and Infancy Development* (ed. Mussen, P.) 281–344 (John Wiley, New York, 1983).
- Kolb, B. Recovery from early cortical damage in rats: I. Differential behavioral and anatomical effects of frontal lesions at 1 or 5 days of age. *Behav. Brain Res.* **25**, 205–220 (1987).
- Goldman, P. S. in *Plasticity and Recovery from Brain Damage* (eds Stein, D. G., Rosen, J. J. & Butters, N.) 149–174 (Academic, New York, 1974).
- Malkova, L., Mishkin, M. & Bachevalier, J. Long-term effects of selective neonatal temporal lobe lesions on learning and memory in monkeys. *Behav. Neurosci.* **109**, 212–226 (1995).
- Mahut, M. & Moss, M. in *The Hippocampus Vol. 4* (eds Isaacson, R. L. & Pribram, K. H.) 241–279 (Plenum, New York, 1986).
- Lipska, B. K. & Weinberger, D. R. Genetic variation in vulnerability to the behavioral effects of neonatal hippocampal damage in rats. *Proc. Natl Acad. Sci. USA* **92**, 8906–8910 (1995).
- Malkova, L., Mishkin, M., Suomi, S. J. & Bachevalier, J. Socioemotional behavior in adult rhesus monkeys after early versus late lesions of the medial temporal lobe. *Ann. NY Acad. Sci.* **807**, 538–540 (1997).
- Fuster, J. M. *The Prefrontal Cortex: Anatomy, Physiology, and Neuropsychology of the Frontal Lobe* (Raven, New York, 1989).
- Roberts, A. C. et al. 6-Hydroxydopamine lesions of the prefrontal cortex in monkeys enhance performance on an analog of the Wisconsin card sort test: possible interaction with subcortical dopamine. *J. Neurosci.* **14**, 2531–2544 (1994).
- Kolachana, B. S., Saunders, R. C. & Weinberger, D. R. Augmentation of prefrontal cortical monoaminergic activity inhibits dopamine release in the caudate nucleus: an *in vivo* neurochemical assessment in the rhesus monkey. *Neuroscience* **69**, 859–868 (1995).
- Moghaddam, B., Berridge, C. W., Goldman-Rakic, P. S., Bunney, B. S. & Roth, R. H. *In vivo* assessment of basal and drug-induced dopamine release in cortical and subcortical regions of the anesthetized primate. *Synapse* **13**, 215–222 (1993).
- Sesack, S. R. & Pickel, V. M. Prefrontal cortical efferents in the rat synapse on unlabeled neuronal targets of catecholamine terminals in the nucleus accumbens septi and on dopamine neurons in the ventral tegmental area. *J. Comp. Neurol.* **320**, 145–160 (1992).
- Taber, M. T., Das, S. & Fibiger, H. C. Cortical regulation of subcortical dopamine release: mediation via the ventral tegmental area. *J. Neurochem.* **65**, 1407–1410 (1995).
- Karremann, M. & Moghaddam, B. The prefrontal cortex regulates the basal release of dopamine in the limbic striatum: an effect mediated by ventral tegmental area. *J. Neurochem.* **66**, 589–598 (1996).
- Kolb, B., Gibb, R. & van der Kooy, D. Neonatal frontal cortical lesions in rats alter cortical structure and connectivity. *Brain Res.* **645**, 85–97 (1994).
- Webster, M. J., Ungerleider, L. G. & Bachevalier, J. Lesions of inferior temporal area TE in infant monkeys alter cortico-amygdalar projections. *Neuroreport* **2**, 769–772 (1991).
- Goldman-Rakic, P. S., Selemon, L. D. & Schwartz, M. L. Dual pathways connecting the dorsolateral prefrontal cortex with the hippocampal formation and parahippocampal cortex in the rhesus monkey. *Neuroscience* **12**, 719–743 (1984).
- Russchen, F. T., Bakst, L., Amaral, D. G. & Price, J. L. The amygdaloid projections in the monkey. An anterograde tracing study. *Brain Res.* **329**, 241–257 (1985).
- Smith, Y. & Parent, A. Differential connections of caudate nucleus and putamen in the squirrel monkey (*Saimiri sciurens*). *Neuroscience* **18**, 347–371 (1986).
- Hikosaka, O., Sakamoto, M. & Miyashita, N. Effects of caudate nucleus stimulation on substantia nigra cell activity in monkey. *Exp. Brain Res.* **95**, 457–472 (1993).
- Bertolino, A. et al. Altered development of prefrontal neurons in rhesus monkeys with neonatal mesial temporo-limbic lesions: a proton magnetic resonance spectroscopic imaging study. *Cerebral Cortex* **7**, 740–748 (1997).
- Jakob, H. & Beckman, H. Prenatal developmental disturbances in the limbic allocortex in schizophrenics. *J. Neural Trans.* **65**, 303–326 (1986).
- Weinberger, D. R. & Berman, K. F. Prefrontal function in schizophrenia: confounds and controversies. *Phil. Trans. R. Soc. Lond. B* **351**, 1495–1503 (1996).
- Carlson, A., Hansson, L. O., Waters, N. & Carlsson, M. L. Neurotransmitter aberration in schizophrenia: new perspectives and therapeutic implications. *Life Sci.* **61**, 75–94 (1997).
- Weinberger, D. R. & Lipska, B. K. Cortical maldevelopment, anti-psychotic drugs, and schizophrenia: a search for common ground. *Schizophrenia Res.* **16**, 87–110 (1995).
- Bebbington, P. E., Bowen, J., Hirsch, S. R. & Kuipers, E. A. in *Schizophrenia* (eds Hirsch, S. R. & Weinberger, D. R.) 587–604 (Blackwell, Oxford, 1995).
- Breier, A. et al. Schizophrenia is associated with elevated amphetamine-induced synaptic dopamine concentrations: evidence from a novel positron emission method. *Proc. Natl Acad. Sci. USA* **94**, 2569–2574 (1997).
- Laruelle, M. et al. Single photon emission computerized tomography imaging of amphetamine-induced dopamine release in drug-free schizophrenic subjects. *Proc. Natl Acad. Sci. USA* **93**, 9235–9240 (1997).
- Kolachana, B. S., Saunders, R. C. & Weinberger, D. R. *In vivo* characterization of extracellular GABA release in the caudate nucleus and prefrontal cortex of the rhesus monkey. *Synapse* **25**, 285–292 (1997).

Acknowledgements. We thank J. Gray for technical assistance and B. Lipska for comments on the manuscript. All procedures were carried out in strict adherence to the NIH Guide for the Care and Use of Laboratory Animals.

Correspondence and requests for materials should be addressed to R.C.S. at his present address (e-mail: rcs@ln.nimh.nih.gov).

Control of spatial orientation in a mollusc

Tatiana G. Deliagina^{*,†}, Yuri I. Arshavsky^{‡,§}
& Grigori N. Orlovsky^{*†}

^{*} The Nobel Institute of Neurophysiology, Department of Neuroscience, Karolinska Institutet, S-171 77, Stockholm, Sweden

[†] AN Belozersky Institute of Physico-Chemical Biology, Moscow State University, Moscow 119 899, Russia

[‡] Institute of Neurobiology, University of Puerto Rico, San Juan, PR00901, USA

[§] Institute of Information Transmission Problems, Russian Academy of Sciences, Moscow 101447, Russia

The main function of postural nervous mechanisms in different species, from mollusc to man, is to counteract the force of gravity and stabilize body orientation in space^{1–3}. Here we investigate the basic principles of postural control in a simple animal model, the marine mollusc *C. limacina*. When swimming, *C. limacina* maintains its vertical orientation because of the activity of the postural neuronal network. Driven by gravity-sensing organs (statocysts), the network causes postural corrections by producing tail flexions. To understand how this function occurs, we studied network activity by using a new method. We used an *in vitro* preparation that consisted of the central nervous system isolated with the statocysts. Output signals from the network (electrical activity of tail motor neurons) controlled an electrical motor which rotated the preparation in space. We analysed the activity of individual neurons involved in postural stabilization under opened or closed feedback loop. When we closed this artificial feedback loop, the network stabilized the vertical orientation of the preparation. This stabilization is based on the tendency of the network to minimize the difference between the activities of the two antagonistic groups of neurons, which are driven by orientation-dependent sensory inputs.

C. limacina (Gastropoda, Opisthobranchia, Pteropoda) is a planktonic animal. Usually it is orientated vertically, with its head

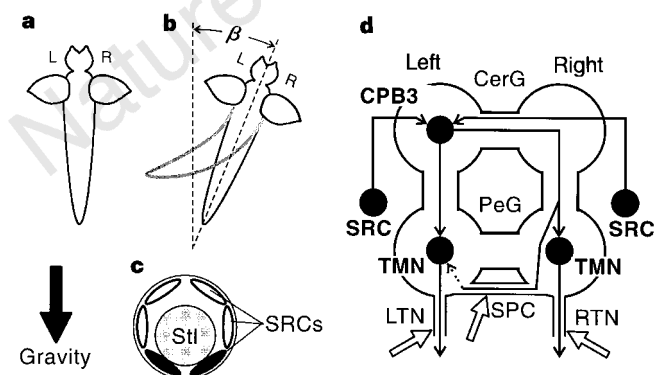


Figure 1 Postural reactions in *C. limacina* and principal elements of the postural network. **a**, Normal orientation of *Clione*. L, left; R, right. **b**, A deviation from this orientation to the right (angle β) produces flexion of the tail to the left. **c**, Structure of the statocyst. The inner cavity of the statocyst is covered by the SRCs. SRCs excited under the effect of the gravity-responding statolith (Stl) are shown in black. **d**, Structure of the cerebral and pedal ganglia (CerG and PeG). We show three neuron classes, which respond in a chain to the gravitational reflexes, and their interconnections. The effects of SRCs from the left and right statocysts on the tail muscles are mediated by the CPB3 interneurons and the TMNs. We recorded gravitational responses of TMNs from their axons in the LTN and RTN; we recorded responses of CPB3 interneurons from their axons in the transected SPC. Recording sites are indicated with open arrows. Responses in SRCs have been studied previously⁷.

up (Fig. 1a). *C. limacina* swims upwards or maintains itself at a particular depth by a continuous beating of two wings^{4–6}. Any deviation from the vertical orientation evokes a corrective motor response (tail flexion) aimed at restoration of the initial orientation⁷. Here we study the nervous control of tail flexion produced by lateral sway, that is, the inclination of the animal in the frontal plane. In Fig. 1b we illustrate the effect of inclination to the right (characterized by the angle β between the longitudinal axis of *C. limacina* and the gravity vector). In response to this inclination, the tail flexes in the opposite direction. The deviated tail, like the rudder of the boat, produces turning of *C. limacina* towards the vertical orientation.

The postural gravitational reflexes are driven by inputs from two statocysts. The statocyst in *C. limacina*, like that in other gastropod molluscs^{8,9}, is spherical (Fig. 1c). About ten statocyst receptor cells (SRCs) cover the inner surface of the statocyst wall¹⁰. A stony-like structure, the statolith, is located in the statocyst cavity. The SRCs are mechanoreceptors which are excited under the effect of the statolith when, because of a change of the organism's orientation, they are found in the lowermost position⁷. After removal of both

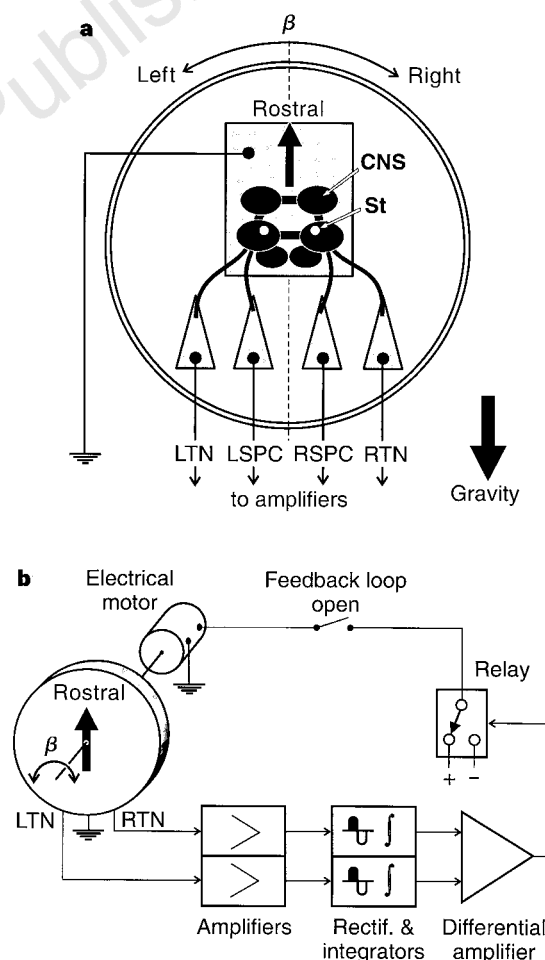


Figure 2 Methods for *in vitro* investigations of the postural network. **a**, Experimental arrangement for gravitational stimulation of the statocysts (St) and for recording the electrical activity from the LTN and RTN and from the left and right stumps of the transected SPC (LSPC and RSPC). Under open-loop conditions, two patterns of rotation of the experimental chamber were produced manually (Figs 3 and 5b, c). The angle of rotation was recorded by a potentiometer transducer. **b**, Electrical circuit for automatic stabilization of the orientation of the CNS in space. The chamber with preparation (shown in **a**) was rotated not manually but by the low-inertia electrical motor controlled by action potentials from the LTN and RTN. For stabilization of CNS orientation, the feedback loop had to be closed.

statocysts, *C. limacina* cannot maintain any definite orientation in space⁷.

Paired recordings of cells in the statocyst and different ganglia of the central nervous system (CNS) have revealed three main cellular groups constituting the postural network, and their interconnections^{7,11} (Fig. 1d). First, there are the SRCs. Second, there is a special class of interneurons (named CPB3 interneurons), which receive excitatory and/or inhibitory inputs from SRCs. These neurons ($n \approx 20$) are located in the cerebral ganglia and vary in their morphology. Roughly six of them send axons into the subpedal commissure (SPC). Here we record activity of these axons in the left and right stumps of the transected commissure. Third, there are several groups of tail motor neurons (TMNs) that are driven by CPB3 interneurons. TMNs elicit flexion of the tail in different planes. Most of these neurons are located in the pedal ganglia. We recorded activity of the axons of tail motor neurons in the left and right lateral tail nerves (LTN and RTN); the axons elicit primarily ipsilateral tail flexion⁷.

The most efficient way to study neuronal networks is to study the isolated CNS or its parts *in vitro*. This method is widely used to analyse, for example, central pattern generators, that is, the neuronal networks capable of rhythmogenesis when isolated^{12–14}. However, isolation of the CNS eliminates the basic function of the postural network, that is, the stabilization of body orientation. To overcome this difficulty, we developed a new method that combines *in vitro* and robotics approaches. We designed a preparation that consists of the CNS and statocysts, thus preserving all basic elements of the postural network. By rotating the preparation in space (see Methods and Fig. 2a), we could investigate the responses of different classes of neurons to natural gravitational stimulation. We then used the output signals of the network, that is, the electrical discharges of tail motor neurons, to control an electrical motor rotating the preparation (Fig. 2b). In this way, we preserved the

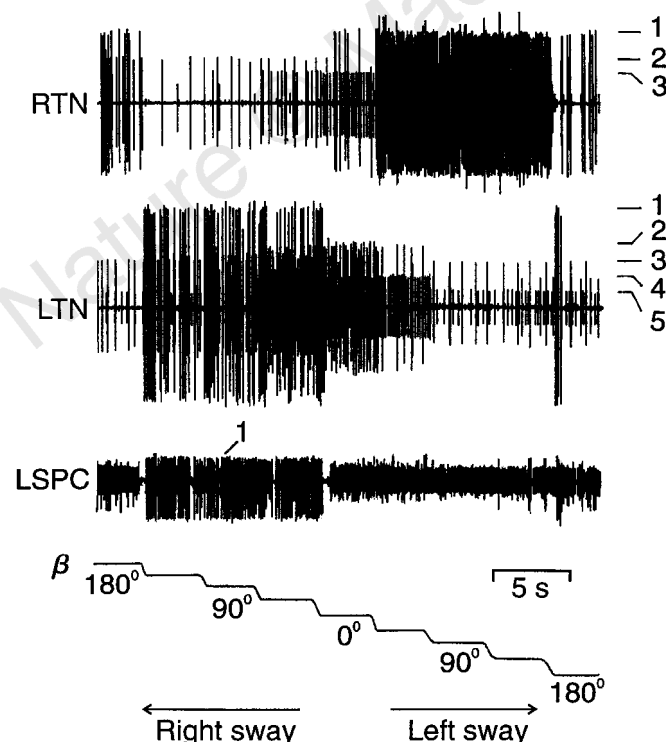


Figure 3 Responses of motor neurons and interneurons to manual rotation of the preparation (opened feedback loop). Motor neurons in the RTN (units 1–3), and in the LTN (units 1–5) were activated preferentially by the contralateral sway. An interneuron with larger spikes, recorded in the left stump of the SPC (LSPC, unit 1), was activated with the right sway. The angle of sway (β) was changed in steps of 45°.

basic function of the network. We investigated activity of the postural network under two conditions, with opened or closed feedback loop.

Figure 3 illustrates responses of motor neurons and interneurons to manual rotation of the preparation through 360°, around its dorsoventral axis; the axis was situated horizontally (Fig. 2a; opened feedback loop). Deviation from the normal, 'head-up' orientation (0°) to the right (right sway) produced activation of motor neurons in the left tail nerve, whereas left sway produced activation of motor neurons in the right tail nerve. Usually, 2–3 (up to 5) motor neurons in each tail nerve responded to gravitational stimulation. The size of the zone of activity of motor neurons varied from 135° to 210° in different experiments ($n = 41$; mean $\pm$ s.d. $166^\circ \pm 36^\circ$). The centre of the zone was found at $64^\circ \pm 23^\circ$ of the contralateral sway.

In each of the stumps of the transected SPC, two CPB3 interneurons responded to the lateral sway (16 experiments), one to the right sway (unit 1 in the trace of the left SPC in Fig. 3), and one to the left sway. The interneurons had similar zones of activity to the corresponding motor neurons: the centre of the zone was situated at $68^\circ \pm 33^\circ$. This result indicates that there may be no essential transformation of the gravitational signals when they pass from interneurons to motor neurons.

Figure 4 shows the effect of closing the feedback loop. When the preparation, moved by the motor (see Methods and Fig. 2b), entered the zone of sensitivity of a particular group of motor neurons (left or right), they became active and produced a reversal in the direction of the motor rotation. Then the preparation entered the zone of the antagonistic group of motor neurons, and the reversal occurred again. Thus, the postural network stabilized the 'head-up' orientation, with periodical oscillations around it. The frequency and amplitude of oscillations depended on the parameters of the feedback loop. With optimal choice of parameters (see Methods), the period of oscillations was 2–4 s and the peak-to-peak value of oscillations (averaged over 35 tests in 11 experiments) was $22^\circ \pm 16^\circ$.

The system with the closed feedback loop seemed to be resistant to externally applied disturbances of the orientation of the preparation in space. As shown in Fig. 4, a large ($\sim 170^\circ$) imposed deviation from the normal orientation, caused by rotation of the preparation when the feedback loop was opened, was rapidly compensated for after the loop was closed again.

C. limacina has been shown to stabilize its 'head-up' orientation only at lower temperatures⁷. At higher temperatures (15–20°C) it stabilizes the 'head-down' orientation and swims downwards, away from warmer water layers⁷. We find that postural activity of *in vitro* preparations ($n = 8$) also depends on temperature (Fig. 5). With the closed feedback loop, the 'head-up' orientation was stabilized at

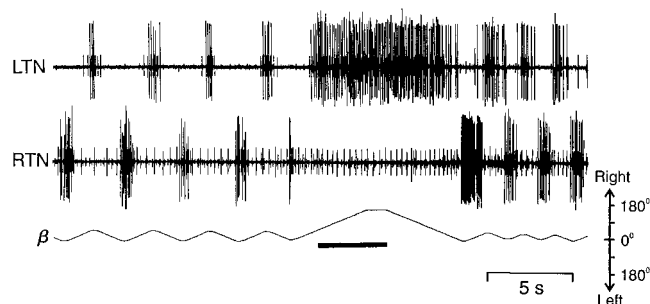


Figure 4 Activity of motor neurons from the LTN and RTN, recorded under the conditions of closed feedback loop. The postural network stabilized the 'head-up' orientation, with small oscillations around it. The thick horizontal bar indicates the period when the feedback loop was opened; we then produced a large deviation from the stabilized orientation by continuously rotating the motor. After the disturbance had terminated and the loop closed again, the system was restored to a normal orientation.

10 °C (Fig. 5a). After the temperature was raised to 20 °C, the system stabilized the 'head-down' orientation (Fig. 5b). A basis for these modifications of postural activity is the reversal of gravitational reflexes, revealed under opened feedback loop conditions (Fig. 5c, d). At 10 °C, the tail motor neurons (in the RTN and LTN) and CPB3 interneuron (in the left SPC) were activated by the contralateral sway (Fig. 5c). At 20 °C, however, they were activated by the ipsilateral sway (Fig. 5d). Thus, temperature-dependent reconfiguration of the postural network occurs at the level of synapses between the SRCs and the interneurons, as shown in Fig. 6. It is unknown, however, whether the synapses themselves are thermosensitive, or whether they are modulated by input from the thermosensitive neurons(s)¹⁵. Thus, the postural network in *C. limacina* is not a hardwired circuit; it can be modified to such an extent that a nearly 'new' network is formed, with a reversed

response to the same gravitational input. Modifications of networks by different modulatory inputs have been described for a number of motor behaviours^{16–18}, including postural control¹⁹.

By combining *in vitro* and robotics approaches, we examined the postural network in *C. limacina* in more detail than is possible with any traditional methods. We characterized the responses to gravitational input in the interneurons and motor neurons (gravitational reflexes), and analysed the activity of the network when it stabilizes the orientation of the preparation in space. Despite the simple design of the 'artificial feedback loop', we have been able to simulate and explain a relatively complex postural activity of the mollusc, including maintenance of different postures. Stabilization of postural orientation is based on the tendency of the network to minimize the difference between the activities of two antagonistic groups of neurons, which are driven by orientation-dependent sensory inputs. This principle seems to apply to systems controlling body orientation in other species, both vertebrate (lamprey²⁰ and rat²¹) and invertebrate (locust^{22,23} and crayfish²⁴). □

Methods

Molluscs were dipped from surface waters in the vicinity of the White Sea biological station Kartesh. Results were obtained from >60 experiments. The experiments were carried out at 10 °C unless indicated otherwise.

Opened feedback loop. Figure 2a shows an experimental arrangement for gravitational stimulation of the statocysts and for recording responses of interneurons and motor neurons. Five electrodes (small pieces of the filter paper soaked in sea water) were fastened to the bottom of the recording chamber. The preparation (CNS isolated with statocysts) was positioned on the indifferent electrode. The LTN and RTN (containing axons of motor neurons eliciting lateral tail flexion) and the left and right parts of the transected SPC (containing axons of CPB3 interneurons) were positioned on the smaller electrodes connected to amplifiers. The chamber was filled with paraffin oil to isolate the electrodes from each other, and then tightly closed. This allowed recording of neuron activity during rotation of the chamber through 360°. Rotation was performed around the horizontally oriented dorsoventral axis of the CNS, which corresponded to the lateral sway of *C. limacina*. The angle β , between the midline of the preparation and the gravity vector, characterized the value of sway.

Closed feedback loop. Figure 2b shows an experimental arrangement for automatic stabilization of the orientation of the preparation in the gravity field. Spikes of motor neuron activity from the LTN and RTN were amplified, rectified and integrated. They were then passed to inputs of the differential amplifier which, through a relay, produced a reverse in the rotation of the electrical motor. If the spike frequency in the left nerve prevailed over that in the right nerve, rotation was to the left; in the opposite case, it was to the right.

When the loop was closed, the system stabilized the 'head-up' orientation with periodical oscillations around it (Fig. 4). The pattern of oscillations depended on the parameters of the feedback loop. An increase of the angular speed of the motor from 10° per s to 60° per s resulted in a proportional increase of the frequency of oscillations, whereas the amplitude of oscillations changed only slightly. In most experiments, the angular speed of the motor was chosen to equal the maximal angular speed (36° per s) that is observed in *C. limacina* during hunting behaviour, when it rotates continuously⁷. An increase of gain in the amplifiers resulted in a decrease of the amplitude and period of oscillations, probably because of the involvement of the motor neurons with smaller spikes in the feedback; these units usually had wider angular zones of sensitivity (Fig. 3). Finally, the time constant of the 'leaking integrator' did not affect the performance practically, provided that it was 2–10 times longer than the average interspike interval in the RTN and LTN. In most experiments, the time constant was 0.1 s.

Received 15 December 1997; accepted 9 March 1998.

- Orlovsky, G. N. Gravistatic postural control in simpler systems. *Curr. Opin. Neurobiol.* **1**, 612–627 (1991).
- Macpherson, J. M., Deliagina, T. G. & Orlovsky, G. N. in *Neurons, Networks, and Motor Behavior* (eds Stein, P. S. G., Grillner, S., Selverston, A. I. & Stuart, D. G.) 257–267 (MIT Press, Cambridge, 1997).
- Horak, F. B. & Macpherson, J. M. in *Handbook of Physiology: Integration of Motor, Circulatory, Respiratory and Metabolic Control during Exercise* (eds Shepard, J. & Rowell, L.) 1–39 (Oxford Univ. Press, 1995).

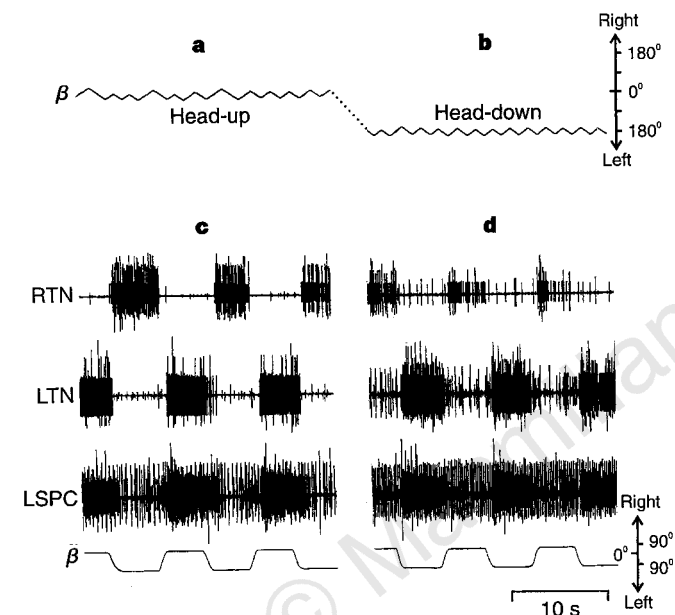


Figure 5 Activity of the postural network depends on temperature. **a, b**, Reversal of the stabilized orientation (closed feedback loop). At 10 °C (**a**) the system stabilized the 'head-up' orientation. With a temperature rise to 20 °C (**b**) the system stabilized the 'head-down' orientation (before this there was a period of instability, indicated by a dotted line). **c, d**, Reversal of gravitational reflexes (opened feedback loop). At 10 °C (**c**), the motor neurons (in the RTN and LTN) and the interneuron (in the left SPC, LSPC) were activated by the contralateral sway. At 20 °C (**d**) they were activated by the ipsilateral sway.

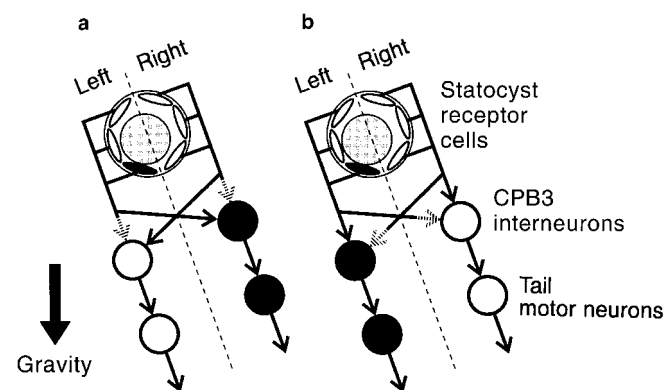


Figure 6 Temperature-dependent reconfiguration of the postural network. Neurons activated by the left sway are shown in black. **a**, At lower temperatures, the SRC excited by gravitational input activates the contralateral reflex chain. **b**, At higher temperatures, the same SRC activates the ipsilateral chain.

4. Arshavsky, Y. I., Beloozerova, I. N., Orlovsky, G. N., Panchin, Y. V. & Pavlova, G. A. Control of locomotion in marine mollusk *Clione limacina*. 1. Efferent activity during actual and fictitious swimming. *Exp. Brain Res.* **58**, 255–262 (1985).
5. Arshavsky, Y. I. *et al.* in *Simpler Nervous Systems* (eds Sakharov, D. A. & Winlow, W.) 290–316 (Manchester Univ. Press, UK, 1991).
6. Satterlie, R. A., LaBarbera, M. & Spencer, A. N. Swimming in the pteropod mollusk *Clione limacina*. 1. Behavior and morphology. *J. Exp. Biol.* **116**, 189–204 (1985).
7. Panchin, Y. V., Arshavsky, Y. I., Deliagina, T. G., Popova, L. B. & Orlovsky, G. N. Control of locomotion of marine mollusk *Clione limacina*. IX. Neuronal mechanisms of spatial orientation. *J. Neurophys.* **73**, 1924–1936 (1995).
8. Alkon, D. L. Responses of hair cells to statocyst rotation. *J. Gen. Physiol.* **66**, 507–530 (1975).
9. Janse, C. Sensory system involved in gravity orientation in pulmonate snail *Lymnaea stagnalis*. *J. Comp. Physiol. A* **145**, 311–319 (1982).
10. Tsiurulis, T. P. The fine structure of the statocyst in the pteropod mollusk *Clione limacina*. *J. Evol. Biochem. Physiol.* **10**, 181–188 (1974).
11. Panchin, Y. V., Arshavsky, Y. I., Deliagina, T. G., Popova, L. B. & Orlovsky, G. N. Control of locomotion in marine mollusk *Clione limacina*. VIII. Cerebropedal neurons. *J. Neurophys.* **73**, 1912–1923 (1995).
12. Arshavsky, Y. I., Deliagina, T. G. & Orlovsky, G. N. Pattern generation. *Curr. Opin. Neurobiol.* **6**, 781–789 (1997).
13. Marder, E. & Calabrese, R. L. Principles of rhythmic motor pattern generation. *Physiol. Rev.* **76**, 687–717 (1996).
14. Selverston, A. I., Panchin, Y. V., Arshavsky, Y. I. & Orlovsky, G. N. in *Neurons, Networks, and Motor Behavior* (eds Stein, P. S. G., Grillner, S., Selverston, A. I. & Stuart, D. G.) 105–117 (MIT Press, Cambridge, 1997).
15. Panchin, Y. V. Cellular mechanisms for the temperature sensitive spatial orientation in *Clione*. *Neuroreport* **8**, 3345–3348 (1997).
16. Harris-Warrick, R. M. & Marder, E. Modulation of neural networks for behavior. *Annu. Rev. Neurosci.* **14**, 39–57 (1991).
17. Selverston, A. I. Neuromodulatory control of rhythmic behaviors in invertebrates. *Int. Rev. Cytol.* **147**, 1–24 (1993).
18. Selverston, A. I. Modulation of circuits underlying rhythmic behaviors. *J. Comp. Physiol. A* **176**, 139–147 (1995).
19. Arshavsky, Y. I. *et al.* Pharmacologically induced elements of the hunting and feeding behavior in the pteropod mollusk *Clione limacina*. 1. Effects of physostigmine. *J. Neurophys.* **69**, 522–532 (1993).
20. Deliagina, T. G. Vestibular compensation in lamprays: impairment and recovery of equilibrium control during locomotion. *J. Exp. Biol.* **200**, 1459–1471 (1997).
21. Deliagina, T. G., Popova, L. B. & Grant, G. The role of tonic vestibular input for postural control in rats. *Arch. Ital. Biol.* **135**, 239–261 (1997).
22. Hensler, K. in *Neurobiology of Sensory Systems* (eds Naresh Singh, R. & Strausfeld, N. J.) 531–554 (Plenum, New York, 1989).
23. Hensler, K. & Rowell, C. H. F. Control of optomotor responses by descending deviation detector neurons in intact flying locust. *J. Exp. Biol.* **149**, 191–205 (1990).
24. Takahata, M., Yoshino, M. & Hisada, M. Neuronal mechanisms underlying crayfish steering behaviour as an equilibrium response. *J. Exp. Biol.* **114**, 599–617 (1985).

Acknowledgements. We thank Ö. Ekeberg, A. Büschges and O. Bloom for comments on the manuscript. This work was supported by the grants from the HHMI (to T.G.D.), Royal Swedish Academy of Sciences (to T.G.D. and G.N.O.) and Swedish Medical Research Council (to T.G.D.).

Correspondence and requests for materials should be addressed to T.G.D. at Karolinska Institutet (e-mail: Tatiana.Deliagina@neuro.ki.se).

Increased vulnerability to cocaine in mice lacking the serotonin-1B receptor

Beatriz A. Rocha^{*†}, Kimberly Searce-Levie^{†‡}, José J. Lucas[§], Noboru Hiroi^{||}, Nathalie Castanon^{‡§}, John C. Crabbe[¶], Eric J. Nestler^{||} & René Hen[‡]

^{*} Department of Pharmacology, University of North Texas Health Science Center, Fort Worth, Texas 76107, USA

[‡] Center for Neurobiology and Behavior, Columbia University, 722 W. 168th Street, P.I. Annex 731, New York, New York 10032, USA

^{||} Laboratory of Molecular Psychiatry, Yale University School of Medicine, New Haven, Connecticut 06508, USA

[¶] Department of Veterans Affairs Medical Center and Oregon Health Sciences University, Portland, Oregon 97201, USA

There is increasing evidence that genetic factors can influence individual differences in vulnerability to drugs of abuse^{1,2}. Serotonin (5-hydroxytryptamine, 5-HT), acting through many receptors can modulate the activity of neural reward pathways and thus the effects of various drugs of abuse^{3–8}. Here we examine

the effects of cocaine in mice lacking one of the serotonin-receptor subtypes, the 5-HT_{1B} receptor⁹. We show that mice lacking 5-HT_{1B} display increased locomotor responses to cocaine and that they are more motivated to self-administer cocaine. We propose that even drug-naïve 5-HT_{1B}-knockout mice are in a behavioural and biochemical state that resembles that of wild-type mice sensitized to cocaine by repeated exposure to the drug. This altered state might be responsible for their increased vulnerability to cocaine.

To assess the reinforcing efficacy of cocaine in 5-HT_{1B}-knockout (KO) mice, we evaluated their ability to self-administer cocaine under a progressive-ratio schedule, in which the number of lever presses required to deliver an injection of cocaine is increased until the mice no longer respond. This 'break point', expressed as the number of injections self-administered, reflects the reinforcing efficacy of the drug tested. Initially, mice were trained to press a lever for food under fixed-ratio 1 or 2 schedules. Both wild-type (WT) and KO mice required ~8 sessions to be trained, indicating a similar ability to learn an operant behaviour. Mice were then implanted with intravenous (i.v.) catheters and trained to self-administer cocaine (2 mg kg⁻¹ per injection; i.v.) under a fixed-ratio-1 schedule. Following acquisition of cocaine self-administration, mice were switched to a progressive-ratio schedule. For all doses of cocaine tested (0.5–4.0 mg kg⁻¹ per injection), KO mice showed significantly higher break points than WT mice (Fig. 1). The KO mice obtained, on average 8 reinforcers (requiring 25 lever presses), whereas the WT mice obtained only 4 (9 lever presses). When saline was substituted for cocaine, both WT and KO mice stopped self-administration (Fig. 1). To confirm the specificity of cocaine's effects, independent groups of mice were trained to respond for food under the progressive-ratio schedule. WT and KO mice obtained an equal number of food reinforcements (Fig. 1). Therefore, the increased break point for cocaine self-administration by the KO does not result from nonspecific hyperactivity in the KO. In summary, cocaine appears to be more reinforcing for KO than WT mice.

The addictive potential of cocaine has been related to its psychomotor stimulant properties as well as its tendency to evoke progressive increases in behavioural response after repeated injections (sensitization)¹⁰. We therefore tested the locomotor responses to cocaine in KO mice after both acute and repeated treatments. When drug-naïve WT and KO mice were placed in an open field

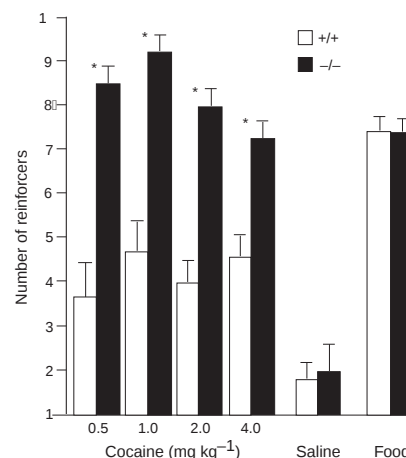
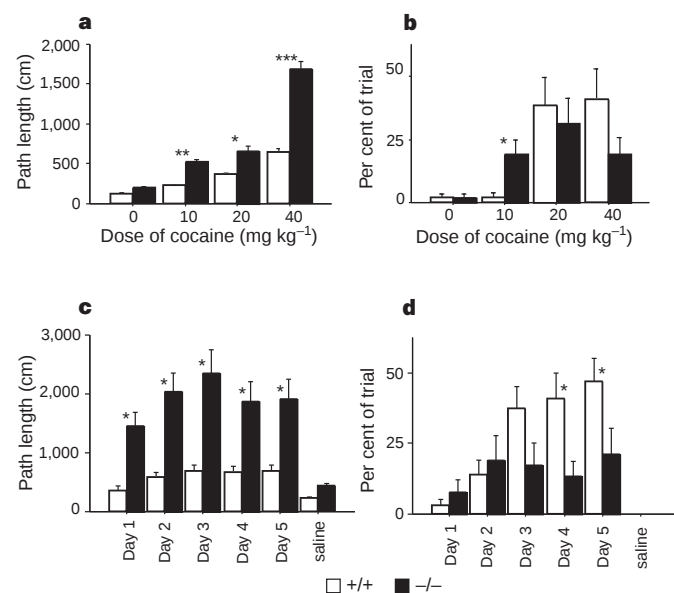


Figure 1 Self-administration behaviour. KO mice (black bars) obtain more cocaine reinforcements, but not food or saline, under a progressive-ratio (PR) reinforcement schedule. Each bar represents the number of reinforcements expressed as mean ± s.e.m. for 10 KO mice and 5 WT (cocaine experiment) or 7 WT and 8 KO (food experiment). Repeated-measures ANOVA confirmed a significant effect of genotype, $P < 0.05$, $F_{(1,13)} = 7.10$. Asterisks indicate a significant ($P < 0.05$) difference between genotypes.

[†] These authors contributed equally to this work.

[§] Present addresses: Centro Biología Molecular Severo Ochoa CBMSO/CSIC, Universidad Autónoma Madrid, Centro Blanco Madrid 28049, Spain (J.L.L.); Laboratory of Molecular Psychobiology, Department of Psychiatry and Neuroscience, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York 10461, USA (N.H.); INSERM, Institute François Magendic, 33077 Bordeaux, France (N.C.).



immediately after a single injection of either saline or cocaine, KO mice showed significantly more locomotion than WT mice in response to 10, 20 and 40 mg kg⁻¹ cocaine (Fig. 2a). In the KO, 10 mg kg⁻¹ cocaine elicited an increase in intense stereotypies (head-weaving and circling in a limited area of the field), whereas only higher doses elicited stereotypy in WT mice (Fig. 2b). For chronic experiments, we studied the effects of five repeated injections of cocaine (20 mg kg⁻¹) on locomotor behaviour. This dose elicited more total locomotion from KO than WT throughout the experiment, although both groups showed increases in locomotion after repeated exposure to cocaine (Fig. 2c). After each cocaine injection, intense stereotypy increased progressively in the WT, indicating behavioural sensitization (Fig. 2d). In the KO, intense stereotypies reached a stable plateau after the second session which was significantly lower than for the WT. These experiments indicate that KO mice tend to respond to cocaine with locomotion, whereas WT are more likely to engage in stereotypy. This behavioural distinction may relate to the fact that the mesostriatal dopamine system contributes more to stereotyped responses to psychostimulants, whereas the mesolimbic system is associated with locomotion and reinforcement¹⁰. The lack of 5-HT_{1B} receptor modulation may have caused the mesolimbic dopamine system to become more dominant in KO mice. Upon first exposure to low doses of cocaine, KO mice showed significantly more locomotion and stereotypies—a pattern of behaviour like that shown by WT mice sensitized by repeated exposure to cocaine.

In normal animals, several biochemical markers are altered after exposure to cocaine. Acute administration of cocaine induces a number of immediate-early genes such as *c-fos* in the dorsal striatum and nucleus accumbens^{11–13}. After repeated cocaine injections, induction of *c-Fos* desensitizes¹³, whereas expression of a set of long-lasting Fos-related proteins, termed chronic FRAs, increases¹⁴. The duration of that change in gene expression, as well as the fact that FRAs are specifically activated after chronic exposure to psychostimulants, indicated that these transcription factors might be involved in the long-lasting effects of drugs of abuse, such as tolerance, sensitization or dependence^{14,15}.

We analysed the expression of FRAs in the brains of WT and KO mice. Striatal protein extracts from drug-naïve mice were analysed by western blotting using an antibody directed against the DNA-binding domain of *c-Fos* that recognizes a family of Fos-related proteins. Several Fos-related proteins, including the 35K–37K chronic FRAs, were more abundant in KO than in WT (Fig. 3a). These chronic FRAs correspond to post-transcriptional modifica-

Figure 2 Effects of cocaine on open field behaviour. Significant differences between KO (black bars) and WT (white bars) are marked by: *, $P < 0.05$; **, $P < 0.01$; or ***, $P < 0.001$. **a, c**, Mean path length + s.e.m. covered in 5 min averaged over 1 h. **b, d**, Mean + s.e.m. of the percentage of time periods stereotypy was exhibited by the mice. **a**, Locomotor response to cocaine. There is a main effect of treatment, $P < 0.0001$, $F_{(3,94)} = 20.72$; a main effect of genotype, $P < 0.0001$, $F_{(1,94)} = 20.61$; and a treatment by genotype interaction, $P < 0.005$, $F_{(3,94)} = 4.90$; $n = 13$ per group. **b**, Stereotypy after acute cocaine. **c**, Locomotion after repeated injections of cocaine (20 mg kg⁻¹). There are main effects of genotype, $P < 0.0001$, $F_{(1,27)} = 24.06$; and day, $P < 0.0005$, $F_{(4,108)} = 6.04$; with no interactions; $n = 16$ per group. **d**, Stereotypy after repeated injections of cocaine. There is a main effect of day, $P < 0.0001$, $F_{(4,108)} = 9.74$; and an interaction of day and genotype, $P < 0.0005$, $F_{(4,108)} = 4.42$.

tions of Δ FosB (a truncated form of FosB that results from alternative splicing of the *fosB* gene)^{16,17}. Although the FosB band is unchanged in KO mice, the Δ FosB bands are more intense in KO than in WT (Fig. 3a, b). We analysed the distribution of the FosB proteins in the striatum by immunocytochemistry on brain sections with an antibody raised against an amino-terminal sequence of FosB that recognizes chronic FRAs^{17,18}. There were more immunoreactive cells in KO than in WT, particularly in the nucleus accumbens (Fig. 3c, d).

Fos-related transcription factors form heterodimers with Jun-related transcription factors. The resulting complex, termed AP-1, binds to a DNA-binding motif found in several promoter regions, and as a result modulates the transcription of downstream genes. To assess directly the presence of AP-1 transcriptional complexes in KO striatum, we did gel retardation assays with labelled oligonucleotides corresponding to the AP-1 binding site. In baseline conditions, the specific bands were more abundant in KO than WT, indicating increased AP-1 transcriptional complex in KO mice (Fig. 4a, b). An anti-FosB antibody was able to induce a supershift of the AP-1 band (Fig. 4c), indicating that most of the AP-1 complex found in KO mice contains FosB.

We also analysed the levels of cocaine and its metabolites, benzoylecgonine and norcocaine, in the brain and blood 15, 30 and 60 min after an acute injection of cocaine (20 mg kg⁻¹)¹⁹. No significant differences were found between genotypes, suggesting that the brain level and rate of cocaine metabolism and elimination is normal in KO mice (results not shown).

Although dopamine, acting through the mesolimbic circuit, is thought to be primarily responsible for cocaine self-administration, 5-HT may modulate this circuit. Animals with 5,7-dihydroxytryptamine lesions of the 5-HT system show an increased breakpoint for cocaine self-administration, suggesting that 5-HT can antagonize the reinforcing effects of cocaine²⁰. However, this type of manipulation, which affects global 5-HT levels, may not reflect the contribution of individual receptor subtypes to the effects of cocaine. 5-HT_{1B} receptors are expressed on the terminals of GABAergic striatal neurons that project to the substantia nigra and ventral tegmental area²¹. Activation of these receptors may inhibit GABA release onto dopaminergic neurons, thereby stimulating the activity of these neurons⁷. 5-HT_{1B} agonists could have cocaine-like activities: the 5-HT_{1B} agonist RU24969 induces increases in locomotor activity⁹ and expression of immediate-early genes such as *c-fos*¹¹; RU24969 can partially substitute for cocaine in drug-discrimination studies⁸ and it potentiates the rewarding effects of cocaine in an intravenous self-

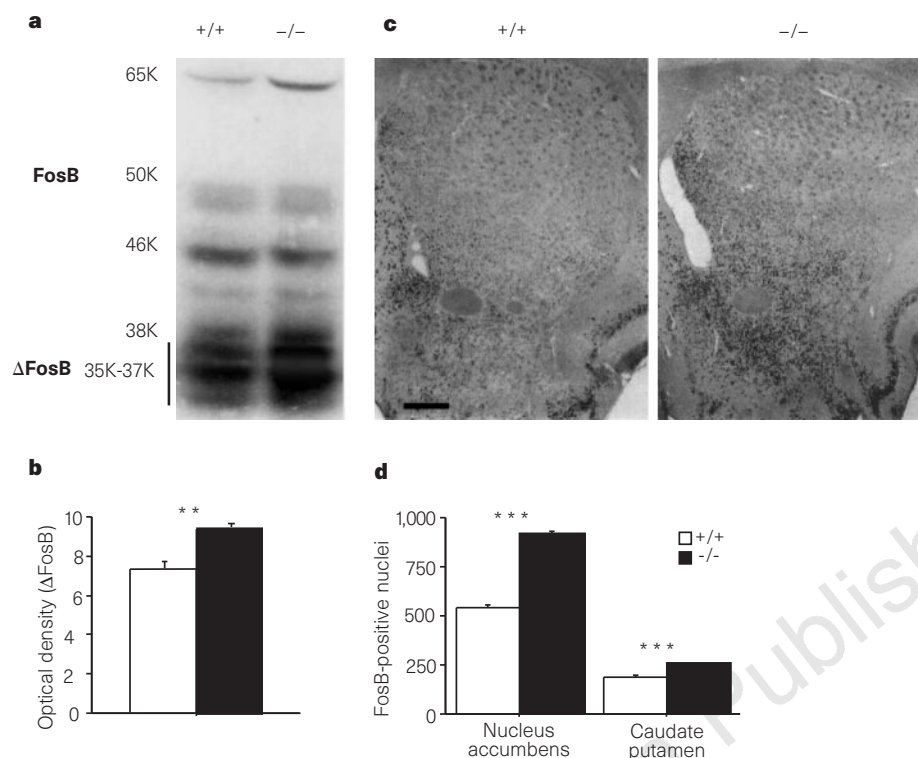


Figure 3 Basal FRAs in WT (+/+) and KO (-/-) mice. **a**, Immunoblot showing ΔFosB bands (corresponding to relative molecular mass 35K–37K) in striatal extracts. Results are representative of five independent experiments. **b**, Optical density shows that the ΔFosB band for 3 KO mice (black bar) is significantly higher ($P < 0.0005$, $F_{(1,4)} = 44.97$) than for 3 WT (white bar). **c**, Immunohistochemistry using a FosB antibody in the nucleus accumbens (NAc) and medial caudate putamen (CPu). Scale bar, 500 μm. **d**, Count of FosB-positive nuclei from 6 WT (white bars) and 6 KO (black bars) mice shows significant genotype differences in NAc, $P < 0.0001$, $F_{(1,10)} = 580.65$; and CPu, $P < 0.0001$, $F_{(1,10)} = 67.76$.

administration protocol²². Therefore a 5-HT_{1B} receptor antagonist should block or decrease the effects of cocaine; we have shown previously that the 5-HT_{1B/1D} antagonist GR127935 decreases the effects of cocaine on c-Fos expression and locomotion^{11,23}. It was therefore surprising that the locomotor response to cocaine in the KO was the opposite of that after antagonist pretreatment. Unlike acute antagonist treatment, the KO could exert its effect throughout the development and life of the mouse, enabling compensatory mechanisms to come into play. The increase in Fos-related transcription factors in KO mice suggests a mechanism that could underlie some of these compensations. This increase seems to be related to an increase in the AP-1 transcription complex, which can regulate the expression of any gene containing AP-1 motifs in its promoter¹⁷. The increased levels of ΔFosB proteins in the nucleus accumbens is particularly interesting because this structure is critical for both the locomotor and rewarding effects of cocaine².

The 5-HT_{1B} KO mice display many of the characteristics of mice that have received a chronic regimen of cocaine and become

sensitized to the drug; increased locomotion; evidence of stereotyped behaviour at lower doses of cocaine; high expression of ΔFosB; reduced induction of c-Fos; and higher breaking points on a progressive-ratio schedule of cocaine self-administration^{11,14,15,24}. As the mesolimbic dopamine system is central to the effects of many categories of drugs of abuse, 5-HT_{1B} KO mice may respond more to other drugs. These mice have already been shown to increase their self-administration of alcohol²⁵. There is evidence that 5-HT_{1B} KO mice are also more impulsive, a trait that is often associated with drug abuse²⁶. 5-HT_{1B} KO mice are therefore a model for some of the biochemical and behavioural changes that might be responsible for individual variations in vulnerability to drugs of abuse. □

Methods

Drugs. Cocaine (Sigma Pharmaceutical) was dissolved in 0.9% saline on the day of the experiment and administered intraperitoneally (i.p.).

Animals. Male mice, 14–20 weeks old, on a pure 129/Sv genetic background, were used for all experiments. Mice were housed 4–5 per cage in a 12-h (06:00–18:00) light/dark colony room at 22 °C with freely available food and water. For the self-administration experiments, mice were housed individually and maintained at 28 g (±5 g) body weight.

Self-administration. Self-administration experiments took place in mouse operant chambers (model ENV-300; Med Associates) as described²⁷. Two days after catheter implantation, mice started cocaine self-administration under a fixed ratio-1 (FR1) schedule (2.0 mg kg⁻¹/0.02 ml⁻¹). At the start of each session, a priming injection of cocaine was given through the catheter. Daily sessions lasted until either 20 injections or 3 h had elapsed. Once mice met acquisition criteria for self-administration (3:1 ratio of active:inactive lever presses; at least 15 injections for 3 consecutive sessions), they were switched to the progressive ratio (PR) schedule. PR sessions also started with the priming injection, but the number of lever presses required to obtain each subsequent injection was based upon the following exponential sequence: 2, 4, 6, 9, 12, 15, 20, 25, 32, ... (ref. 28). Sessions lasted for three hours or until mice failed to respond for one hour. Once mice reached a stable baseline of responding (number of injections not varying by more than 20% over three consecutive days), different doses of cocaine (0.5, 1.0 and 4.0 mg kg⁻¹ per injection) were tested. Each dose was tested on three consecutive days in a random order. The

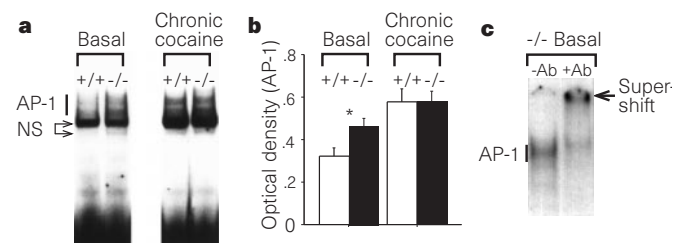


Figure 4 Basal AP-1 DNA-binding activity in striatal samples. **a**, Autoradiogram of a gel mobility-shift assay showing AP-1 DNA-binding activity in striatal extracts after 8 daily injections of 20 mg kg⁻¹ cocaine or saline. NS, nonspecific binding. Results are representative of five independent experiments. **b**, Optical density of AP-1 bands for 3 KO and 3 WT mice (basal) and 6 KO and 5 WT mice (chronic cocaine). Genotypes differ significantly in control conditions, $P = 0.05$, $F_{(1,4)} = 7.55$. **c**, Autoradiogram of supershift assay. Left lane, no FosB antibody; right lane, after addition of FosB antibody. Results are representative of six independent experiments.

number of injections earned was used as the dependent variable for statistical analysis (univariate and repeated-measures ANOVA).

Open field locomotion. Testing was conducted between 08:00 and 17:00. Animals were placed in 40-cm² square, open-field chambers. They were monitored throughout each 1-h test session by a video-tracking system (Poly-Track, San Diego Instruments) which records the animal's location and path. The mean path length for each 5-min interval was calculated and analysed using one-way repeated-measures ANOVA. Post-hoc Scheffé tests were used to compare conditions. Animals were also videotaped throughout all test sessions for later evaluation of stereotyped behaviour. On the two days preceding the acute cocaine experiment, animals were given a saline injection, then placed in the open field to allow them to habituate to the chamber. On the third day, animals were given an injection of either saline or cocaine (10, 20 or 40 mg kg⁻¹) and monitored in the open field. For the chronic cocaine experiment, all animals received five total injections of cocaine (20 mg kg⁻¹), with a 2-day rest between each injection. At the end of the repeated cocaine regimen, animals were given a challenge injection of saline and monitored.

Behavioural scoring. Two trained observers, who were blind to both mouse genotype and drug treatment, evaluated stereotyped behaviour. Each animal was observed for 1 min at four different time points during the test session: 2, 15, 30 and 45 minutes post-injection.

Immunohistochemistry. Brain sections were prepared and treated as described¹¹. Sections were incubated overnight at 4°C in affinity-purified primary antibody raised against amino acids 102–117 of the mouse *fosB* gene (FosB[102] sc-48; Santa-Cruz Biotechnology) diluted 1:10,000 (ref. 14). Omission of the primary antibody or preadsorption with the control peptide (sc-48P) resulted in no labelling. Immunopositive nuclei were counted as described¹¹.

Isolation of brain regions for western blot and gel shift. Following decapitation, striatum was excised and extracts were prepared as described¹⁴. For supershift assay, striatal nuclear extracts were used.

Gel shift assay. The gel shift and supershift assays were done as described¹⁷. The AP-1 probe was a double-stranded synthetic oligonucleotide derived from the promoter sequence of the human metallothionein II gene²⁹. Electrophoresis and blotting conditions are described elsewhere¹⁴. Bands were shown to be specific for AP-1 binding by cold competition and supershift assays as before¹⁴.

Western blotting. Samples were prepared and electrophoresed as described¹⁴. Blots were incubated with an anti-M peptide primary antibody (anti-FRA, 1:4,000; from M. Iadarola) followed by horseradish peroxidase-conjugated goat anti-rabbit IgG (1:4,000). Quantification of signals was by densitometry. The specificity of the resulting immunoreactive bands was confirmed by abolishing immunoreactivity when anti-FRA antibody was preadsorbed with purified M peptide¹⁴.

Received 27 January; accepted 2 March 1998.

- Crabbe, J. C., Belknap, J. K. & Buck, K. J. Genetic animal models of alcohol and drug abuse. *Science* **264**, 1715–1723 (1994).
- Koob, G. F. Drugs of abuse: anatomy, pharmacology and function of reward pathways. *Trends Pharmacol. Sci.* **13**, 177–184 (1992).
- Parsons, L. H., Weiss, F. & Koob, G. F. Serotonin1b receptor stimulation enhances dopamine-mediated reinforcement. *Psychopharmacology* **128**, 150–160 (1996).
- Peltier, R. & Schenk, S. Effects of serotonergic manipulations on cocaine self-administration in rats. *Psychopharmacology* **110**, 390–394 (1993).
- Carroll, M. E., Lac, S. T., Ascencio, M. & Kragh, R. Fluoxetine reduces intravenous cocaine self-administration in rats. *Pharmacol. Biochem. Behav.* **35**, 237–244 (1990).
- Parsons, L. H. & Justice, J. B. Jr. Perfusion of serotonin increases extracellular dopamine in the nucleus accumbens as measured by *in vivo* microdialysis. *Brain Res.* **606**, 195–199 (1993).
- Cameron, D. L. & Williams, J. T. Cocaine inhibits GABA release in the VTA through endogenous 5-HT. *J. Neurosci.* **14**, 6763–6767 (1994).
- Callahan, P. M. & Cunningham, K. A. Modulation of the discriminative stimulus properties of cocaine by 5-HT1B and 5-HT2C receptors. *J. Pharmacol. Exp. Ther.* **274**, 1414–1424 (1995).
- Saudou, F. *et al.* Enhanced aggressive behavior in mice lacking 5-HT1B receptor. *Science* **265**, 1875–1878 (1994).
- Wise, R. & Bozarth, M. A psychomotor theory of addiction. *Psychol. Rev.* **94**, 469–492 (1987).
- Lucas, J. J., Segu, L. & Hen, R. 5-Hydroxytryptamine1B receptors modulate the effect of cocaine on c-fos expression: converging evidence using 5-hydroxytryptamine1B knockout mice and the 5-hydroxytryptamine1B/1D antagonist GR127935. *Mol. Pharmacol.* **51**, 755–763 (1997).
- Graybiel, A. M., Moratalla, R. & Robertson, H. A. Amphetamine and cocaine induce drug-specific activation of the c-fos gene in striosome-matrix compartments and limbic subdivisions of the striatum. *Proc. Natl Acad. Sci. USA* **87**, 6912–6916 (1990).
- Young, S. T., Porrino, L. J. & Iadarola, M. J. Cocaine induces striatal c-fos-immunoreactive proteins via dopaminergic D1 receptors. *Proc. Natl Acad. Sci. USA* **88**, 1291–1295 (1991).
- Hope, B. T. *et al.* Induction of a long-lasting AP-1 complex composed of altered Fos-like proteins in brain by chronic cocaine and other chronic treatments. *Neuron* **13**, 1235–1244 (1994).
- Nestler, E. J., Hope, B. T. & Widnell, K. L. Drug addiction: a model for the molecular basis of neural plasticity. *Neuron* **11**, 995–1006 (1993).

- Chen, J., Kelz, M. B., Hope, B. T., Nakabeppu, Y. & Nestler, E. J. Chronic Fos-related antigens: stable variants of deltaFosB induced in brain by chronic treatments. *J. Neurosci.* **17**, 4933–4941 (1997).
- Hiroi, N. *et al.* FosB mutant mice: loss of chronic cocaine induction of Fos-related proteins and heightened sensitivity to cocaine's psychomotor and rewarding effects. *Proc. Natl Acad. Sci. USA* **94**, 10397–10402 (1997).
- Mandelzys, A., Gruda, M. A., Bravo, R. & Morgan, J. I. Absence of a persistently elevated 36-kDa fos-related antigen and AP-1 like DNA binding activity in the brains of kainic acid-treated fosB null mice. *J. Neurosci.* **17**, 5407–5415 (1997).
- Browne, S. P., Moore, C. M., Scheuren, J., Tebbett, I. R. & Logan, B. K. A rapid method for the determination of cocaine in brain tissue. *J. Forens. Sci.* **36**, 1662–1665 (1991).
- Loh, E. A. & Roberts, D. C. Break-points on a progressive ratio schedule reinforced by intravenous cocaine increase following depletion of forebrain serotonin. *Psychopharmacology* **101**, 262–266 (1990).
- Boschert, U., Amara, D. A., Segu, L. & Hen, R. The mouse 5-hydroxytryptamine1B receptor is localized predominantly on axon terminals. *Neuroscience* **58**, 167–182 (1994).
- Parsons, L. H., Koob, G. F. & Weiss, F. Serotonin-1B receptor stimulation enhances cocaine reinforcement, cocaine-induced increases in nucleus accumbens dopamine, and cocaine-induced decreases in ventral tegmental area GABA. (Soc. Neurosci. 26th Annu. Meeting, Washington, DC, 1996).
- Castanon, N., Lucas, J. J., Searce, K. & Hen, R. The 5-HT1B/1D antagonist GR 127935 decreases the effects of cocaine on c-fos expression and locomotion. (Soc. Neurosci. 26th Annu. Meeting, Washington, DC, 1996).
- Horger, B. A., Shelton, K. & Schenk, S. Preexposure sensitizes rats to the rewarding effects of cocaine. *Pharmacol. Biochem. Behav.* **37**, 707–711 (1990).
- Crabbe, J. C. *et al.* Elevated alcohol consumption in null mutant mice lacking 5-HT1B serotonin receptors. *Nature Genet.* **14**, 98–101 (1996).
- Brunner, D. & Hen, H. In *The Neurobiology of Suicide: From the Bench to the Clinic* (eds Stoff, D. M. & Mann, J. J.) 81–105 (N.Y. Acad. Sci., New York, 1997).
- Rocha, B., Ator, R., Emmett-Oglesby, M. & Hen, R. Intravenous cocaine self-administration in mice lacking 5-HT1B receptors. *Pharmacol. Biochem. Behav.* **57**, 407–412 (1997).
- Depoortere, R. Y., Li, D. H., Lane, J. D. & Emmett-Oglesby, M. W. Parameters of self-administration of cocaine in rats under a progressive-ratio schedule. *Pharmacol. Biochem. Behav.* **45**, 539–548 (1993).
- Sonnenberg, J. L., Macgregor-Leon, P. F., Curran, T. & Morgan, J. I. Dynamic alterations occur to the levels and composition of transcription factor AP-1 complexes after seizure. *Neuron* **3**, 359–365 (1989).

Acknowledgements. We thank E. Gardner and J. Chen for biochemical characterization of the mice, R. Aton for help with self-administration studies, and R. Yarmolinsky for help with preparation of the figures. This work was supported by grants from NIDA, Bristol-Meyers, and a HHMI predoctoral fellowship (K.S.-L.).

Correspondence and requests for materials should be addressed to R.H. (e-mail: rh95@columbia.edu).

Frizzled signalling controls orientation of asymmetric sense organ precursor cell divisions in *Drosophila*

Michel Gho & François Schweisguth

Laboratoire de Génétique du Développement de la *Drosophile*, URA 1857, Ecole Normale Supérieure, 46 rue d'Ulm, 75005 Paris, France

During metazoan development, cell-fate diversity is brought about, in part, by asymmetric cell divisions¹. In *Drosophila*, bristle mechanosensory organs are composed of four different cells that originate from a single precursor cell, pI, after two rounds of asymmetric division². At each division, distinct fates are conferred on sister cells by the asymmetric segregation of Numb, a negative regulator of Notch signalling^{3–6}. Here we show that the orientation of the mitotic spindles and the localization of the Numb crescent follow a stereotyped pattern. Mitosis of pI is orientated parallel to the anteroposterior axis of the fly. We show that signalling mediated by the Frizzled receptor polarizes pI along this axis, thereby specifying the orientation of the mitotic spindle and positioning the Numb crescent. The mitoses of the two cells produced by mitosis of pI are orientated parallel and orthogonal, respectively, to the division axis of pI. This difference in cell-division orientation is largely independent of the identity of the secondary precursor cells, and is regulated by Frizzled-independent mechanisms.

We analysed the orientation of pI divisions in the dorsocentral region of the pupal notum. Mitotic spindles were aligned parallel to the anteroposterior body axis (Fig. 1a, b) ($10 \pm 18^\circ$ relative to the anteroposterior axis; $n = 48$). Moreover, in all cases, Numb was distributed in an anterior crescent, and was segregated to the anterior daughter cell. The posterior daughter cell, pIIa, divides first to generate the shaft and socket cells, and the anterior daughter

cell, pIIb, divides next to give rise to the neuron and the sheath cell (Fig. 1g).

Mitotic spindles of dividing pIIa were also aligned parallel to the anteroposterior axis (Fig. 1c, d) ($7 \pm 28^\circ$; $n = 44$). In all cases ($n = 44$), Numb accumulated in the anterior pole of pIIa, next to pIIb, and was segregated to the anterior daughter, which differentiated into a shaft cell⁷. These data are consistent with the role of Numb in cell-fate specification of pIIb and of shaft cell^{3,8}.

pIIb divides soon after pIIa. Mitotic spindles of pIIb were orientated roughly orthogonal to the previous pIIa division axis ($80 \pm 22^\circ$, $n = 19$; Fig. 1e, f). Numb was localized at the lateral pole, that is, away from the midline, of the dividing pIIb ($n = 17$ out of 19). As the lateral daughter cell inherits Numb, we predicted that the

lateral cell adopts a neuronal fate. Indeed, the neuron, identified 24 h after pupal formation (APF) using antibodies against horse-radish peroxidase, was most frequently found at the anterolateral position ($n = 18$ out of 22; not shown). Conversely, the sheath cell, detected at 23 h APF using anti-Prospero antibodies, was found in the anteromedial position ($n = 35$ out of 39; Fig. 1g).

These results show that each asymmetric division in the bristle lineage is spatially orientated. Hence, each sense-organ cell occupies a fixed position relative to each other and to the fly body axis (Fig. 1g). This pattern of orientated mitosis will probably be essential for sensory functions, as short sequences of orientated mitosis have been observed in sensillum lineages in various arthropod species^{9,10}. Finally, the position of the neuron and sheath cell relative to the midline revealed the existence of a previously unknown mediolateral polarity in the notum.

Because of the lack of cell-specific markers, the identity of pIIa and pIIb is defined by the fate of their daughter cells. Our results show that pIIa and pIIb also differ in their axes of division, which are orthogonal to each other. This immediately raises the possibility that the orientation of division of secondary precursors depends on cell identity. To test this hypothesis, pIIb was transformed into a second pIIa by ectopically activating Notch signalling using an activated form of the receptor, *Nintra*¹¹ (Fig. 2a). Transient overexpression of *Nintra* at 16.5 h APF, before pI division, resulted in the formation of sensory organs composed of two shaft and two socket cells (Fig. 2b, c). These organs resulted from the transformation of pIIb into pIIa. The orientation of division of the ectopic pIIa was inferred from the relative position of the cells within clusters composed of two socket and two shaft cells at 24 h APF. In 85% of these cases ($n = 35$ out of 42), the ectopic, anteriorly located pIIa divided with an orientation similar to that of non-transformed pIIb (Fig. 2d, f; angle values $>22.5^\circ$). We conclude that the pIIb-to-pIIa cell-fate transformation is not sufficient to change the orientation of cell division of ectopic pIIa. In the remaining cases ($n = 7$ out of 42), the four sister cells were aligned along the anteroposterior axis (Fig. 2e, f; angle values $<22.5^\circ$), indicating that the ectopic pIIa cell divided with an orientation identical to its pIIa sister. Such cell alignments were never seen in sibling pupae, indicating that they resulted from the action of *Nintra*. Thus Notch signalling may regulate not only cell fates but also aspects of cell polarity at mitosis.

The stereotyped pattern of mitosis in the bristle lineage (Fig. 1g) indicates that tissue polarity may regulate the orientation of these divisions. Mutations in the *frizzled* (*fz*) and *dishevelled* (*dsh*) genes affect tissue polarity and, in particular, disrupt bristle orientation on the notum¹². *Fz* is the putative receptor for an as yet unidentified polarity signal¹³. *Dsh* is a signal-transducing protein that functions downstream of *Fz*¹⁴. Mutations in *dsh* (*dsh*¹) and *fz* (*fz*⁵⁴) resulted in randomly orientated pI divisions in the epithelial plane (Fig. 3a, g). The Numb crescent also localized in a random manner. However, its position still tightly correlated with the position of one pole of the misorientated spindle. We conclude that *Fz*-mediated signalling provides polarity information to pI, specifying both the orientation of mitotic spindles and the position of the Numb crescent. Mitotic spindles were never orientated perpendicular to the epithelial plane in *fz* or *dsh* mutant pupae. This indicates that a *Fz*-independent mechanism may restrict the positioning of mitotic spindles to this plane.

Only pI cells responded to *Fz*/*Dsh* signalling. The orientation of pIIa division was examined in *dsh*¹, *fz*^{K21}/*fz*^{KD4a} (a transheterozygous combination of null *fz* alleles¹⁵) and *fz*⁵⁴ mutant pupae. In these mutants, the socket, shaft and pIIb cells were always aligned in a three-cell row, as in wild-type pupae (Fig. 3b, d–f, h). Furthermore, in dividing pIIa cells, the Numb crescent was always localized in the region of the cell in contact with pIIb (Fig. 3b). Consistent with this, the socket cell always occupied an eccentric position (Fig. 3d–f). This indicates that, despite the lack of tissue polarity, the orientation of the pIIa division relative to the position of pIIb remains properly

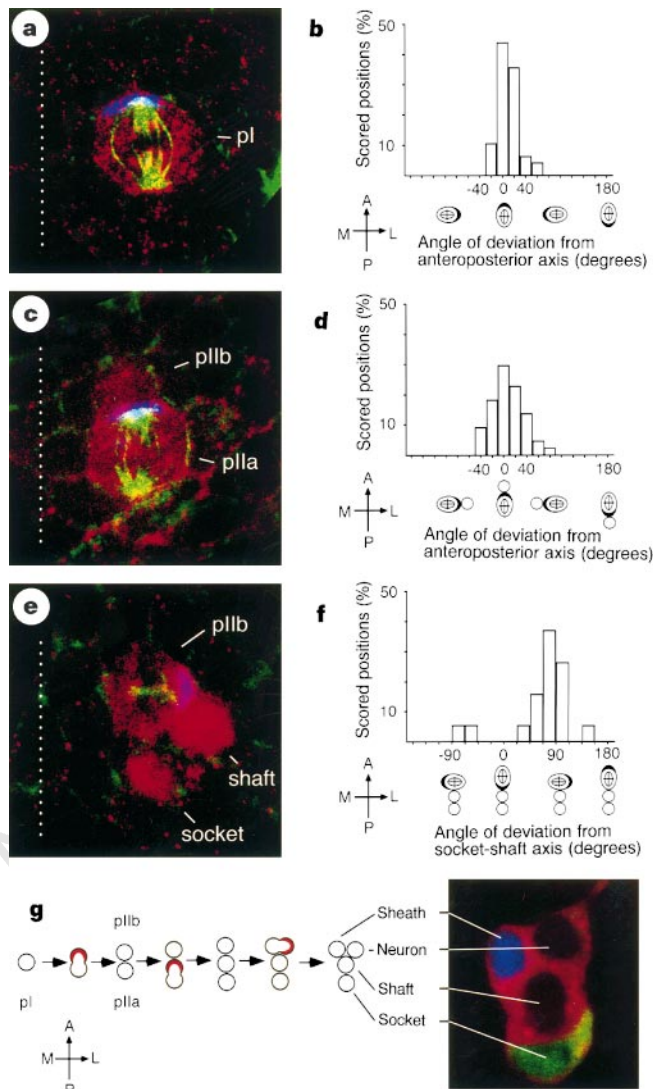


Figure 1 Orientation of cell divisions in the bristle lineage. **a, c, e**, Confocal images of nota from A101 pupae showing pI (**a**), pIIa (**c**) and pIIb (**e**) divisions. β -Galactosidase (nuclear staining) in red, α -tubulin in green and Numb in blue. **b, d, f**, Histograms showing the orientation of pI, pIIa and pIIb divisions, respectively. Spindle orientation (abscissa; see diagrammatic representation below each histogram) was measured relative to the midline (**b, d**) or to the axis joining the two pIIa daughter cells (**f**). Ordinate results are given as a percentage of the total number of positions analysed ($n = 48, 44$ and 19 for **b, d** and **f**, respectively). **g**, Left, diagram of the bristle lineage; Numb is shown in red. Right, a sensory-organ cluster of a dA-10 pupa at 23 h APF. Suppressor of hairless (*Su(H)*), a socket-cell marker⁷, is shown in green, β -galactosidase in red and Prospero in blue. In this and the following figures, anterior is up. The midline is denoted by a dashed white line in **a, c, e**. A, anterior; P, posterior; M, medial; L, lateral.

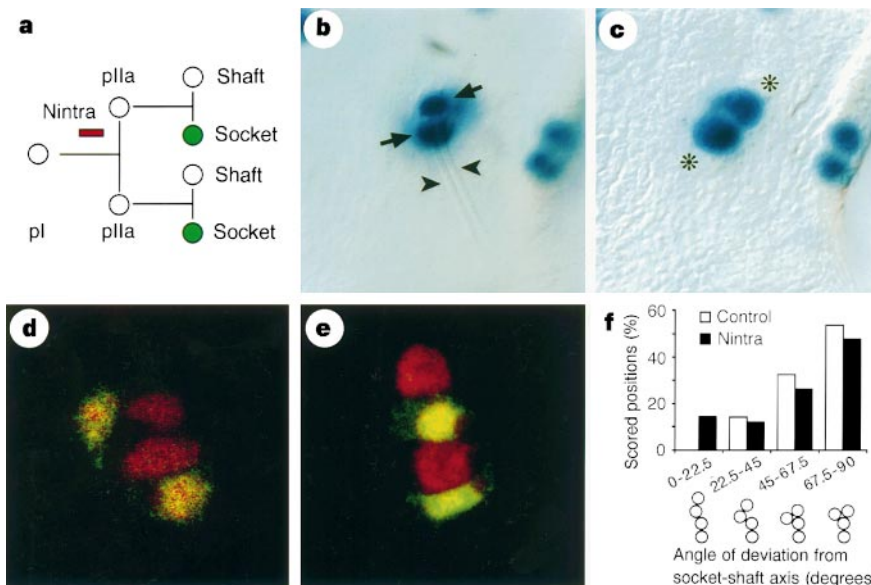


Figure 2 The orientation of the pllb division is largely independent of its cell identity. **a**, Schematic view of the bristle lineage after pllb-to-plla cell-fate transformation induced by expression of Nintra just before pl division. Sensory organs resulting from such a transformation were composed of two pairs of socket and shaft cells. **b**, Epidermal focal plane showing the two shafts (arrowheads) and the two socket-cell nuclei (arrows; β -galactosidase activity staining of A101 pupae at 40 h APF). **c**, The plla subepidermal focal plane showing the two shaft nuclei (asterisks). **d**, **e**, Relative position of the socket and shaft cell pairs following pllb-to-plla transformation. A101 pupae at 24 h APF; β -galactosi-

dase is shown in red, and Su(H) in green. In most cases, the two pairs were orthogonally located (**d**). In some cases, all cells were aligned along the anteroposterior axis (**e**). **f**, Histogram showing the angle between the axis joining the two nuclei of the ectopic, anteriorly located, pair of socket and shaft cells and the axis joining the two endogenous, posteriorly located, socket and shaft cells. Black bars, hs-Nintra/+; A101/+ heat-shocked pupae ($n = 42$); white bars: +/Cyto; A101/+ heat-shocked sibling pupae ($n = 43$). Angle values around 0° were never observed in control pupae.

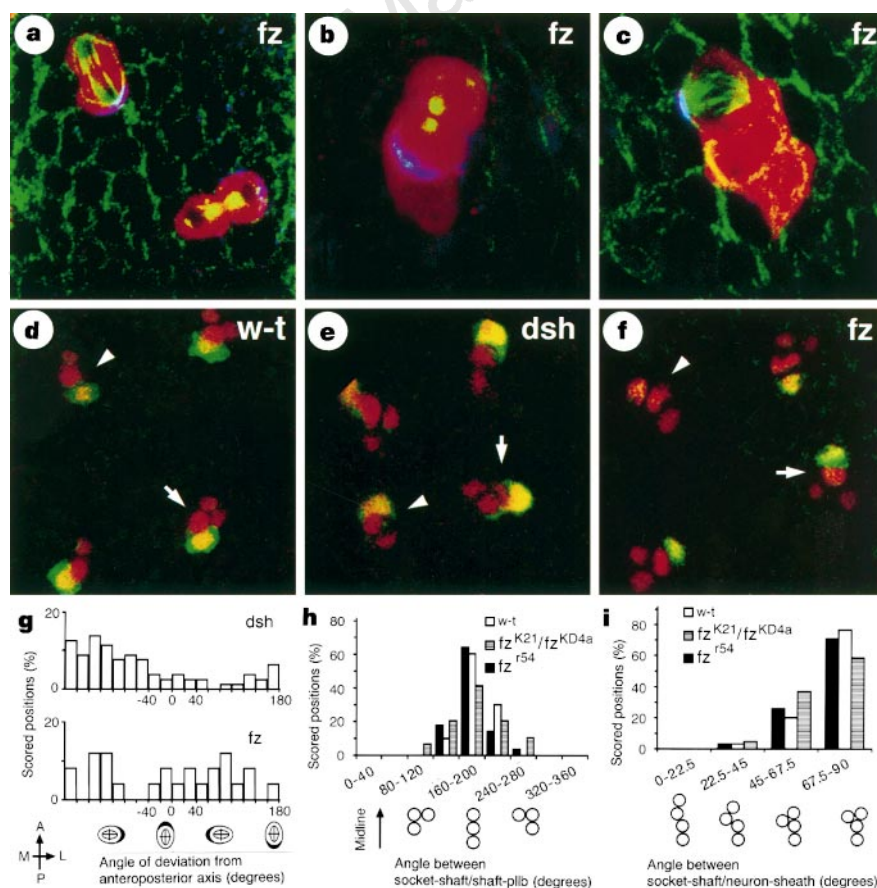


Figure 3 Orientation of cell divisions in *fz* and *dsh* mutant pupae. Confocal images of nota from dA-10; *fz*⁵⁴ (**a-c**), A101 (**d**), *dsh*¹; A101 (**e**) and *fz*^{K21}/*fz*^{KD4a} (**f**) flies. β -Galactosidase (cytoplasmic in **a-c**; nuclear in **d, e**) and Cut (**f**) are shown in red, staining all sensory-organ cells; α -tubulin (**a-c**) and Su(H) (**d-f**) are shown in green; and Numb is in blue (**a-c**). **a-c**, pl, plla and pllb divisions in *fz* mutant pupae, respectively. **d-e**, Like wild-type pupae (**d**), three-cell rows (arrowheads) and four-cell clusters forming perpendicular pairs of cells (arrows) were seen in *dsh*¹ (**e**) and *fz*^{K21}/*fz*^{KD4a} (**f**) mutant nota at 24 h APF. Socket cells (in green) were always located eccentrically. **g**, Histograms showing the orientation of pl divisions in *fz*⁵⁴ ($n = 25$) and *dsh*¹ ($n = 79$) mutant pupae. In *dsh* mutant pupae, the orientation of the pl division was inferred from the orientation of the three-cell rows. **h**, Histogram showing the angle between the socket-shaft and the shaft-pllb cell pairs in wild-type (w-t) pupae (dA-10; $n = 20$) and *fz*^{K21}/*fz*^{KD4a} ($n = 29$) and *fz*⁵⁴ ($n = 28$) mutant pupae. **i**, Histogram showing the angle between socket-shaft and neuron-sheath cell pairs in wild-type pupae (dA-10; $n = 43$) and *fz*^{K21}/*fz*^{KD4a} ($n = 41$) and *fz*⁵⁴ ($n = 42$) mutant pupae. A, anterior; P, posterior; M, medial; L, lateral.

defined in these mutants. Similarly, disruption of Fz signalling did not perturb the orthogonal orientation of pIIb division and Numb localization relative to its sister cells (Fig. 3c). Thus, characteristic perpendicular pairs of cell clusters were observed in *fz* and *dsh* mutant, as in wild-type, pupae (Fig. 3d–f, i).

These data show that the orientation of the pIIa and pIIb divisions is not determined by Fz-mediated tissue polarity. They further indicate that the axis of pI division and/or the relative position of its daughter cells regulates the orientation of the pIIa and pIIb divisions. Two observations are consistent with this conclusion. First, the variability in angle values for pIIa division was larger than that seen for pI, as if it resulted from the addition of its own variability and those observed for pI (Fig. 1b, d). Second, the variability in angle values for pIIb division is lower when angles were measured relative to pIIa daughter cells ($80 \pm 22^\circ$; Fig. 1f) than when they were measured relative to the notum midline ($75 \pm 35^\circ$).

Our results indicate that the polarity of the three mitotic cells in the bristle lineage are regulated by distinct mechanisms. The polarity of pI is regulated by Fz signalling. Fz seems to specifically regulate the orientation of pI divisions, as epidermal cell divisions were not spatially orientated in the notum (results not shown). In contrast to pI, the orientations of the pIIa and pIIb divisions are independent of Fz signalling and, instead, are defined by the axis of the previous pI division. This could be either through positioned cortical marks, as described for yeast¹⁶, or through cell–cell signalling between sister cells.

Studies of neuroblast divisions in the embryo indicate that Inscuteable (Insc) coordinates both the orientation of mitotic spindles and the position of the Numb crescent¹⁷. As the mitotic spindle and the Numb crescent are both misorientated in a coordinated manner in *fz* and *dsh* mutant pI cells, we propose that Fz signalling controls the localization and/or the activity of an organizer acting similarly to Insc in pI. Insc is unlikely to be such a pI organizer because clonal analysis shows that *insc* regulates neither bristle differentiation nor polarity in the notum (not shown). Similarly, in the embryo, *insc* activity appears to be dispensable in mitotic sense-organ cells that have a planar polarity¹⁷. Insc may thus act primarily in apicobasal polarity in epithelial cell. The organizer acting downstream of Fz signalling in planar divisions remains to be identified.

Control of cell polarity by Fz signalling has been evolutionarily conserved, at least from nematodes to flies^{18–20}. However, the *mom-5* and *lin-17* genes, which encode Fz-like receptors in *Caenorhabditis elegans*, are required to establish both orientation of division and asymmetry, whereas *fz* is required to regulate the orientation, but not the asymmetry, of pI divisions. Whether Fz signalling also contributes to determining cell polarity during asymmetric divisions in vertebrates is unknown. □

Methods

Drosophila stocks. The A101 line carries an *P(lacZ, ry⁺)* enhancer-trap allele of *neuralised* that specifically expresses nuclear β -galactosidase in pI and its progeny cells²¹. The dA-10 transformant line carries a *deadpan-lacZ* construct expressing cytoplasmic β -galactosidase in pI and its progeny cells²² (gift from E. Bier). The following mutant alleles were used: *dsh¹*, *fz^{K21}*, *fz^{K21}* and *fz^{KD4a}* (gifts from F. Chanut, D. Coen and P. Adler). Lineage analysis was conducted in A101/TM6b, dA-10, *fz^{K21}/fz^{KD4a}*, dA-10; *fz^{K21}*, *dsh¹*; A101/+. The hs-Nintra/CyO transformant line I4a (gift from T. Lieber) was used to overexpress an activated form of Notch (for 30 min at 37°C).

Immunostaining. Dissected nota from pupae at 16–24 h APF were processed as described⁷. Primary antibodies were rabbit anti-Numb (gift from Y.-N. Jan; 1:1000); mouse anti- β -galactosidase (Promega; 1:2000); rat anti- α -tubulin (Serotec; 1:3000); rat anti-Su(H) (1:500), mouse anti-Prospero (MR1A, gift from C. Doe; 1/2), mouse anti-Cut (2B10, DSHB; 1:500) and rabbit anti- β -galactosidase (Cappel; 1:500). Secondary conjugated antibodies were: Indodicarbocyanine-anti-rabbit (Jackson; 1:500); Indodicarbocyanine-anti-mouse (Jackson; 1:500); fluorescein-isothiocyanate-anti-rat (Jackson; 1:100); liss-

amine rhodamine sulphonyl chloride-anti-mouse (Jackson; 1:200); tetramethyl rhodamine isothiocyanate anti-rabbit (Biosys; 1:200). Confocal images were obtained on a Leica TCS 4D confocal microscope and images were processed with NIH images and Photoshop programs.

Received 19 November 1997; accepted 10 March 1998.

- Horvitz, H. & Herskowitz, I. Mechanisms of asymmetric cell division: two Bs or not two Bs, that is the question. *Cell* **68**, 237–255 (1992).
- Hartenstein, V. & Posakony, J. W. Development of adult sensilla on the wing and notum of *Drosophila melanogaster*. *Development* **107**, 398–405 (1989).
- Rhyu, M. S., Jan, L. Y. & Jan, Y. N. Asymmetric distribution of Numb protein during division of the sensory organ precursor cell confers distinct fates to daughter cells. *Cell* **76**, 477–491 (1994).
- Knoblich, J. A., Jan, L. Y. & Jan, Y. N. Asymmetric segregation of Numb and Prospero during cell division. *Nature* **377**, 624–627 (1995).
- Frise, E., Knoblich, J. A., Younger-Shepherd, S., Jan, L. Y. & Jan, Y. N. The *Drosophila* Numb protein inhibits signaling of the Notch receptor during cell–cell interaction in sensory organ lineage. *Proc. Natl Acad. Sci. USA* **93**, 11925–11932 (1996).
- Guo, M., Jan, L. Y. & Jan, Y. N. Control of daughter cell fates during asymmetric division: interaction of Numb and Notch. *Neuron* **17**, 27–41 (1996).
- Gho, M., Lecourtis, M., Géraud, G., Posakony, J. W. & Schweisguth, F. Subcellular localization of Suppressor of Hairless in *Drosophila* sensory organ cells during Notch signalling. *Development* **122**, 1673–1682 (1996).
- Wang, S., Younger-Shepherd, S., Jan, L. Y. & Jan, Y. N. Only a subset of binary cell fate decisions mediated by Numb/Notch signaling in *Drosophila* sensory organ lineage requires *Suppressor of Hairless*. *Development* **124**, 4435–4446 (1997).
- Lawrence, P. A. Gradients in the insect segment: the orientation of hairs in the milkweed bug *Oncopeltus fasciatus*. *J. Exp. Biol.* **44**, 607–620 (1996).
- Bate, C. M. in *Handbook of Sensory Physiology* (ed. Jakobson, M.) 1–52 (Springer, Berlin, 1978).
- Lieber, T., Kidd, S., Alcano, E., Corbin, V. & Young, M. W. Antineurogenic phenotypes induced by truncated Notch proteins indicate a role in signal transduction and may point to a novel function for Notch in nuclei. *Genes Dev.* **7**, 1949–1965 (1993).
- Adler, P. N. The genetic control of tissue polarity in *Drosophila*. *Bioessays* **14**, 735–741 (1992).
- Park, W.-J., Lin, J. & Adler, P. N. The *frizzled* gene of *Drosophila* encodes a membrane protein with an odd number of transmembrane domains. *Mech. Dev.* **45**, 127–137 (1994).
- Krasnow, R. E., Wong, L. L. & Adler, P. N. *dishevelled* is a component of the *frizzled* signaling pathway in *Drosophila*. *Development* **121**, 4095–4102 (1996).
- Jones, K. H., Lin, J. & Adler, P. N. Molecular analysis of EMS-induced *frizzled* mutations in *Drosophila melanogaster*. *Genetics* **142**, 205–215 (1996).
- Gonczy, P. & Hyman, A. Cortical domains and the mechanisms of asymmetric cell division. *Trends Cell Biol.* **6**, 382–387 (1996).
- Kraut, R., Chia, W., Jan, L. Y., Jan, Y. N. & Knoblich, J. A. Role of *inscutable* in orientating asymmetric cell divisions in *Drosophila*. *Nature* **383**, 50–55 (1996).
- Thorpe, C. J., Schlesinger, A., Clayton Carter, J. & Bowerman, B. Wnt signaling polarizes an early *C. elegans* blastomere to distinguish endoderm from mesoderm. *Cell* **90**, 695–705 (1997).
- Rocheleau, C. et al. Wnt signaling and an APC-related gene specify endoderm in early *C. elegans* embryos. *Cell* **90**, 707–716 (1997).
- Sawa, H., Lobel, L. & Horvitz, H. R. The *Caenorhabditis elegans* lin-17, which is required for certain asymmetric cell division, encodes a putative seven-transmembrane protein similar to the *Drosophila* Frizzled protein. *Genes Dev.* **10**, 2189–2197 (1996).
- Usui, K. & Kimura, K. Sequential emergence of the evenly spaced microchaetes on the notum of *Drosophila*. *Arch. Dev. Biol.* **203**, 151–158 (1993).
- Emery, J. F. & Bier, E. Specificity of CNS and PNS regulatory subelements comprising pan-neural enhancers of the *deadpan* and *scratch* genes is achieved by repression. *Development* **121**, 3549–3560 (1995).

Acknowledgements. We thank P. Adler, E. Bier, D. Coen, F. Chanut, C. Doe, Y.-N. Jan, T. Lieber and the Developmental Studies Hybridoma Bank for flies and antibodies; F. Rosa and M. Wassef for critical reading of the manuscript; and the Imaging facility of the Institut Jacques Monod for use of confocal microscopes. This work was supported by grants from the CNRS, the MENESR, the Fondation pour la Recherche Médicale, the Association pour la Recherche contre le Cancer and the Ligue Nationale contre le Cancer—Comité de Paris.

Correspondence and requests for materials should be addressed to M.G. (e-mail: mgcho@wotan.ens.fr).

The molecular elasticity of the extracellular matrix protein tenascin

Andres F. Oberhauser, Piotr E. Marszalek, Harold P. Erickson* & Julio M. Fernandez

Department of Physiology and Biophysics, Mayo Foundation, Rochester, Minnesota 55905, USA

* Department of Cell Biology, Duke University Medical Center, Durham, North Carolina 27710, USA

Extracellular matrix proteins are thought to provide a rigid mechanical anchor that supports and guides migrating and rolling cells^{1–4}. Here we examine the mechanical properties of the extracellular matrix protein tenascin by using atomic-force-microscopy techniques. Our results indicate that tenascin is an

elastic protein. Single molecules of tenascin could be stretched to several times their resting length. Force–extension curves showed a saw-tooth pattern, with peaks of force at 137 pN. These peaks were ~25 nm apart. Similar results have been obtained by study of titin⁵. We also found similar results by studying recombinant tenascin fragments encompassing the 15 fibronectin type III domains of tenascin. This indicates that the extensibility of tenascin may be due to the stretch-induced unfolding of its fibronectin type III domains. Refolding of tenascin after stretching, observed when the force was reduced to near zero, showed a double-exponential recovery with time constants of 42 domains refolded per second and 0.5 domains per second. The former speed of refolding is more than twice as fast as any previously reported speed of refolding of a fibronectin type III domain^{6,7}. We suggest that the extensibility of the modular fibronectin type III region may be important in allowing tenascin–ligand bonds to persist over long extensions. These properties of fibronectin type III modules may be of widespread use in extracellular proteins containing such domain^{8,9}.

Tenascins are extracellular matrix proteins that are conserved in all vertebrates^{10,11}. Their functional roles include cell adhesion and the consequent mechanical interactions between cells. These proteins are typically disulphide-linked hexamers called hexabrachions. Each tenascin subunit is made up of a series of repeated structural domains, including tandem epidermal growth factor (EGF)-like repeats, tandem fibronectin type III (FN-III) domains and a terminal knob region homologous to a region of fibrinogen^{10,11}. Tenascin mediates the attachment and rolling of cells in shear flow¹². The attachment site on tenascin is the terminal knob, which binds to an unknown receptor on the rolling cells. A significant feature of tenascin is that the splice variants differ only in the number of FN-III repeats (8–15), indicating that the tandem FN-III modules may have a functional role that is finely regulated^{10,11}.

The FN-III domain is a common motif of modular proteins and is found in ~2% of all animal proteins¹³. The FN-III domain is characterized by its folding into a seven-stranded β -barrel, a structure that is very similar to that of immunoglobulin domains^{8,14}. The tandem repeats of immunoglobulin domains of the giant

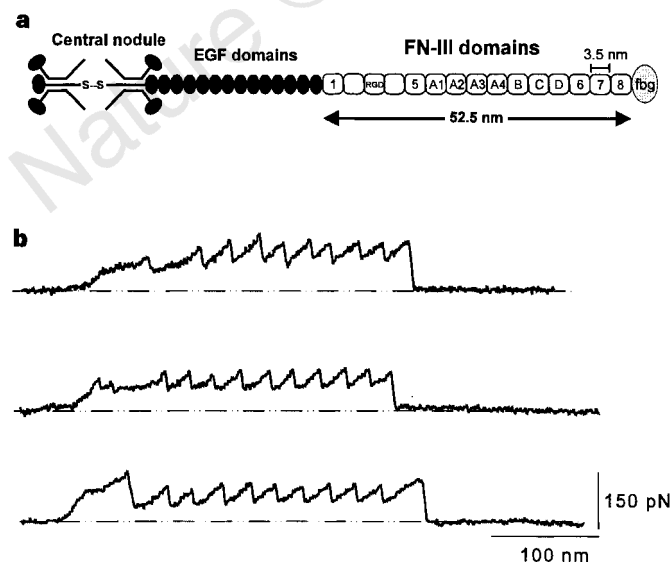


Figure 1 Force–extension relationships for native tenascin hexabrachions measured with AFM techniques. **a**, Modular construction of one subunit of the hexameric tenascin^{10,11}. The EGF domains of tenascin have internal disulphide bonds (S–S) that prevent extension¹¹. Extension of tenascin involves unfolding FN-III domains, and perhaps part of the central knob. (RGD, single-letter amino-acid code; fbg, fibrinogen.) **b**, Stretching single hexabrachions gave force–extension curves that followed a saw-tooth pattern with equally spaced force peaks.

muscle protein titin function as an extensible shock absorber with hysteresis^{5,15–17}. The elastic properties of the tandem immunoglobulin region of titin are due to the forced unfolding of individual immunoglobulin domains. In the atomic-force-microscope (AFM) experiments, rapid unfolding was seen when the immunoglobulin domains were strained by forces of up to 300 pN (ref. 5). It seemed likely that the tandem FN-III domains of tenascin could also act as an extensible segment in the mechanical interactions of tenascin with cells and extracellular matrix.

We have used AFM techniques^{5,18–20} to examine the mechanical properties of the human tenascin-C protein (Fig. 1). We used the large splice variant of tenascin-C which has 15 FN-III domains per subunit²¹. The protein was adsorbed onto a gold-coated coverslip and placed in the recording chamber of a single-axis AFM. Random segments of tenascin-C were picked up by an AFM tip and then stretched. The resulting force–extension curves showed a saw-tooth pattern, with as many as 12 peaks with an average force of 137 ± 12 pN ($n = 42$; pulling speed, $0.2\text{--}0.6$ nm ms^{−1}). The force peaks were equally spaced by a distance of 24.8 ± 2.3 nm ($n = 37$). A unique feature of the tenascin force–extension curves is a ‘hump’ that precedes the first force peak. Incubating tenascin with the

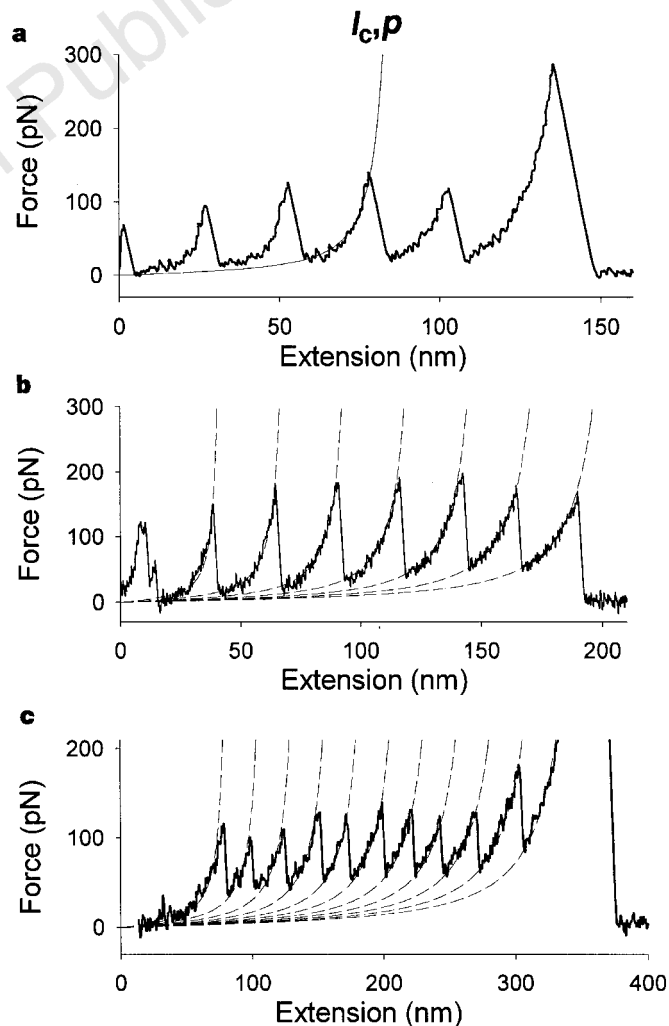


Figure 2 The WLC model, using a persistence length of $p = 0.42$ nm and a contour length increment of $\Delta l_c = 28.5$ nm, describes the force–extension curves of recombinant tenascin fragments. **a**, Force–extension curve for a TNfnA-D protein. The smooth curve corresponds to the fit of an individual force–extension peak to the WLC equation $F(x) = kT/p(x/l_c + 1/4(1 - x/l_c)^2 - 1/4)$. **b**, **c**, A family of WLC curves generated with the parameters found in **a** describes the force–extension curves of the TNfnALL protein (**b**; dotted lines) and of native dithiothreitol-treated tenascin (**c**; dotted lines).

reducing reagent dithiothreitol (1 mM for 1 h, a treatment that disrupts the hexabrachion into single arms; H.P.E., unpublished data) eliminated the hump. This hump may therefore result from stretching the α -helical and central knob domains of the native hexabrachion.

We also studied the recombinant tenascin segment TNfnALL²¹, which is composed of 15 FN-III domains and corresponds to the largest splice variant, and TNfnA-D²¹, which is made of seven FN-III domains and corresponds to one of the spliced segments in native tenascin. Stretching the recombinant proteins TNfnALL and TNfnA-D produced saw-tooth-like force-extension curves (Fig. 2).

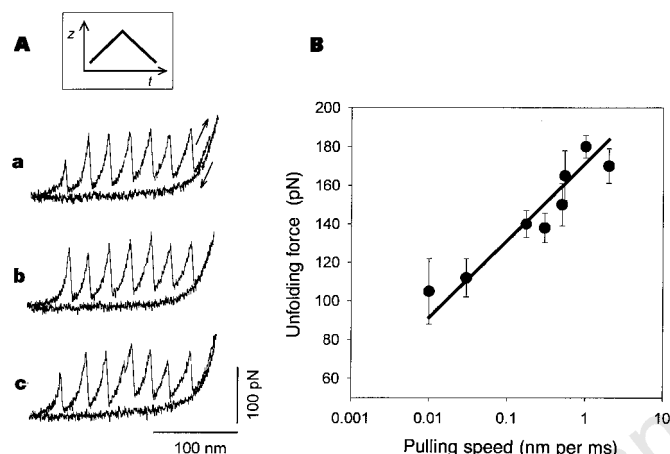


Figure 3 Repeated unfolding/refolding cycles of a single recombinant TNfnALL protein. **A**, Three consecutive unfolding/refolding cycles separated by 5 s (**a**, **b**) and 15 s (**b**, **c**). **a**, Upwards arrow indicates extension and downwards arrow indicates relaxation. **B**, Dependence of the unfolding forces on pulling speed. The filled circles show the result of five different experiments ($n = 191$). The solid line is the result of a Monte Carlo simulation of the speed dependency of the force peaks measured during the simulated unfolding of ten FN-III domains placed in series. The inset shows z -axis motion as a function of time.

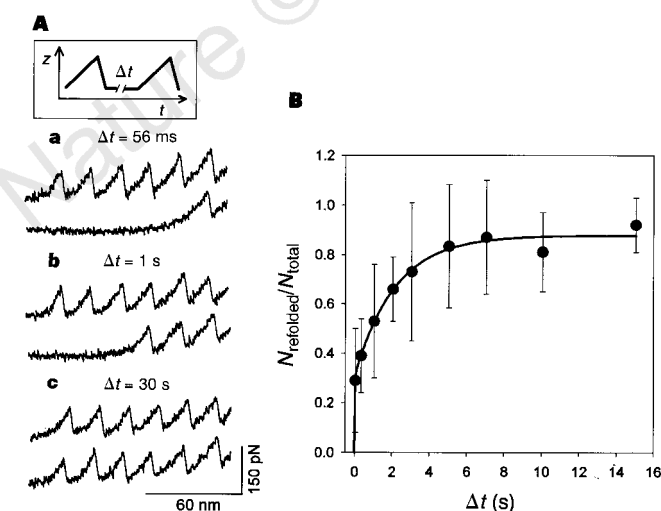


Figure 4 Repeated refolding cycles of a single TNfnAll protein using a double-pulse experiment (inset) identifies at least two refolding rate constants. **A**, A first extension identifies the total number of domains that unfold, N_{total} . (**a-c**, top traces). Then, the protein is relaxed to its resting length to allow refolding. After a delay, Δt , a second extension measures the number of refolded domains, N_{refolded} . (**a-c**, bottom traces). **B**, Plot of $N_{\text{refolded}}/N_{\text{total}}$ versus Δt . Each symbol is the average of (from left to right) 79, 40, 45, 22, 12, 3, 14 and 6 data points obtained from 4 separate experiments. The solid line is a two-exponential fit of the data to the function $N_{\text{refolded}}/N_{\text{total}} = A_1(1 - e^{-\Delta t/\beta_{01}}) + A_2(1 - e^{-\Delta t/\beta_{02}})$, where $A_1 = 0.3$, $A_2 = 0.57$, $\beta_{01} = 42 \text{ s}^{-1}$ and $\beta_{02} = 0.5 \text{ s}^{-1}$.

The force peaks of TNfnALL and TNfnA-D had an average force of $138 \pm 50 \text{ pN}$ ($n = 146$; pulling speed, $0.3\text{--}0.5 \text{ nm ms}^{-1}$) and a spacing of $24.7 \pm 1.2 \text{ nm}$.

The elasticity of a polypeptide chain is well described by the worm-like-chain (WLC) model, which predicts the relationship between the extension of a polymer and the entropic restoring force generated^{5,16,17}. The adjustable parameters of the WLC model are the persistence length, p , and the contour length of the polymer, l_c (see ref. 22 for a complete description). Nonlinear fits (Levenberg-Marquardt fits) of the WLC function to the force-extension curves that lead to each force peak showed that the WLC model adequately describes the elasticity of stretched tenascin (Fig. 2a, solid line). The fits gave an average persistence length of $p = 0.42 \pm 0.22 \text{ nm}$ ($n = 75$). The contour length determined by the fits ranged from 25 nm to 496 nm. However, the change in contour length between consecutive force peaks was a constant, $\Delta l_c = 28.5 \pm 4.0 \text{ nm}$ ($n = 96$). A family of WLC curves generated with a persistence length of 0.42 nm and contour length increments of 28.5 nm describes well the force-extension curves of the recombinant tenascin protein TNfnALL (Fig. 2b) and of native tenascin (Fig. 2c).

The size of a folded tenascin FN-III domain was estimated from the crystal structure to be 3.6 nm (ref. 23). The tenascin FN-III domains consist of between 88 and 92 amino acids. When the

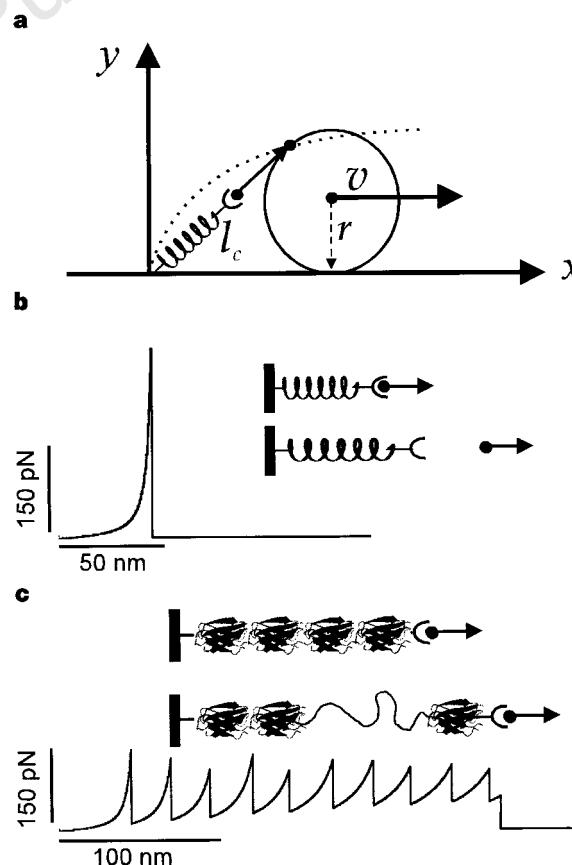


Figure 5 A Monte Carlo simulation shows that tandem FN-III repeats can extend the range and lifetime of a protein-ligand bond and reduce the force required to break it (see Methods). **a**, During cell rolling, a protein bound to a cell ligand is stretched along a cycloidal path (dotted line). The protein is represented by a coil and the ligand by a black circle near to the coil. **b**, Simulation of a bound 50-nm-long protein, without extensible modules, stretched at a speed of 0.2 nm ms^{-1} , predicts a force-extension curve where bond breakage occurs after a 45-nm extension and at a force of 330 pN. **c**, At the same pulling speed, the simulated force-extension curve of a cell-bound protein composed of 15 FN-III domains (inset; only four are shown) shows bond breakage after an extension of 268 nm and at 69 pN.

polypeptide is maximally extended, each amino acid contributes 0.38 nm to the contour length, so unfolding a domain could lengthen the subunit by 31 nm. This is slightly more than the observed increase of 28.5 nm and implies that extension to 90% of maximum requires a force sufficient to unfold the next domain.

We could stretch and relax a protein repeatedly if we limited the extension so that the molecule did not detach from the AFM tip or the substrate. Under these conditions, we observed that extension always showed the saw-tooth pattern of unfolding, whereas relaxation showed a monotonic force–relaxation curve (Fig. 3A). Repeated extension–relaxation curves showed a similar pattern, showing that unfolding was fully reversible. Refolding could not be observed during the force–relaxation curve, indicating that domain refolding may not have begun until the protein was relaxed to, or near to, its original length (Fig. 3A). There should be some small force associated with the refolding, but it is too small to measure with our instrument.

The rate of extension of a tenascin protein is an important factor determining the peak forces at which unfolding was observed. We examined this by varying the pulling speed over a wide range (0.01–2 nm ms⁻¹). A 100-fold reduction in the pulling speed reduced the unfolding force by ~75 pN (Fig. 3B). The speed dependence of the unfolding force is well described by a Monte Carlo simulation that assumes a two-state model for the folding/unfolding reaction (Fig. 3B, solid line). In this simulation, we used the unfolding length that was found for the titin immunoglobulin modules ($\Delta x_u = 0.3$ nm) (ref. 5) and the unfolding rate constant determined for the third FN-III domain of tenascin ($\alpha_u = 4.6 \times 10^{-4}$ s⁻¹) (ref. 24).

Refolding of the TNfnALL protein was not instantaneous: it was important to wait several seconds between consecutive extensions for the saw-tooth pattern to be fully restored. To study this in detail, we used a double-pulse protocol where the pulse interval was varied (Fig. 4A, B). The number of FN-III domains refolded, after a complete unfolding, recovered with a double-exponential time course. Roughly 30% of the recovery occurred at a fast rate (42 s⁻¹) whereas 57% occurred at a much slower rate (0.5 s⁻¹). As we were counting the number of domains completely renatured (that is, domains capable of generating a stretch peak), these results indicate that some domains refold at the fast rate and some at the slower rate. Double-exponential refolding of FN-III domains has been observed with circular dichroism and fluorescence techniques^{6,7,24}. However, the fast refolding rates observed here are faster than those observed in chemical renaturing. For example, a value of 2.9 s⁻¹ has been found for TNfn3 (ref. 24), and values of 20 and 0.04 s⁻¹ have been found for the major, fastest folding of the FN10 and FN9 subunits, respectively, from fibronectin. Therefore, some of the domains of tenascin are refolding at more than twice the fastest rate previously reported for an FN-III domain, which were already surprisingly fast considering the presence of multiple prolines⁶. The AFM is therefore a powerful tool for studying refolding of single protein domains at a rapid timescale.

Mechanical interactions between cells involve receptor–ligand bonds, which can be challenged by forces in the range 100–200 pN during rolling adhesions^{3,4}. In these reactions, the membrane receptors and/or their matrix ligands are typically elongated, multi-domain proteins. Tenascin has been shown to bind to an as yet unknown receptor on lymphocytes to support cell rolling¹². Tenascin-dependent cell rolling is likely to trigger folding/unfolding reactions of the FN-III modular region of tenascin. The dynamic extensibility of this region could allow the tenascin–receptor bond to persist over long extensions; because the bond persists over a longer time the eventual rupture force will be lower (Fig. 5). These properties of FN-III modules may be used in other proteins containing such domains. □

Methods

Single-molecule atomic-force microscopy. Our atomic-force microscope

was constructed using a Digital Instruments (Santa Barbara, SA, USA) AFM detector head (AFM-689) mounted on top of a Physik Instrumente (Waldbronn, Germany) single-axis piezoelectric positioner (P732.ZC). Data sampling and control of the piezoelectric positioner were done by means of an AT-MIO-16X data-acquisition board driven by LabView software (National Instruments). The spring constant of each individual AFM cantilever was calibrated in solution using the equipartition theorem as described²⁵. This method gives values for the spring constant of the cantilever which are within 20% of the values obtained by other methods²⁵. Human native and recombinant tenascin-C proteins were expressed and purified as described²¹. The proteins were suspended in phosphate-buffered saline at a concentration of 10–100 µg ml⁻¹ and were allowed to adsorb onto freshly evaporated gold coverslips. Force–extension curves for single tenascin proteins were obtained under conditions similar to those used to stretch single titin proteins⁵.

Monte Carlo simulations of domain unfolding. The unfolding of a FN-III domain was represented by a two-state Markovian model with force-dependent rate constants^{5,17}. This description gives the probability of observing the unfolding of any module as $P_u = N_f \alpha \Delta t$, where N_f is the number of folded modules, α is the unfolding rate constant and Δt is the polling interval. Similarly, the folding probability was calculated as $P_f = N_u \beta \Delta t$, where N_u is the number of unfolded modules and β is the folding rate constant. The effect of an external force on the unfolding and refolding rate constants was calculated as $\alpha = \alpha_0 \exp(F \Delta x_u / kT)$ and $\beta = \beta_0 \exp(-F \Delta x_f / kT)$, where F is the applied force and $\Delta x_u = 0.3$ nm and $\Delta x_f = 1$ nm are the unfolding and folding distances, respectively. The rates in the absence of an applied force, $\alpha_0 = 4.6 \times 10^{-4}$ s⁻¹ and $\beta_0 = 3$ s⁻¹, were taken from ref. 24. However, as refolding is not observed in the presence of a force (Fig. 3A), these values of β_0 and Δx_f have no impact in the simulation of a single force–extension curve. However, they become significant when simulating unfolding/refolding cycles (see below).

In order to simulate the extension of a tenascin protein, the force experienced by the FN-III modules during a ramp extension was calculated by using the WLC model with parameters identical to those obtained in Fig. 2a ($p = 0.42$ nm; $\Delta l_c = 28.5$ nm). Once the force for a given extension was determined, the probability of unfolding of any module was calculated and then each module was polled to determine its status, following the Monte Carlo approach. Using these methods we predicted the force–extension curve for a TNfnALL protein at various rates of extension and measured the average value of the resulting force peaks. Data from several simulations were pooled together and interpolated, giving the solid line shown in Fig. 3B.

Simulations of the stretching of a protein–ligand bond during cell rolling.

We simulated protein–ligand-dependent cell rolling (cell radius $r = 5$ µm; rolling velocity $v = 1$ µm s⁻¹) where a cell-bound protein is stretched along a cycloidal path (Fig. 5a, dotted line). At this rolling speed, a cell-bound protein extended by ~100 nm in length will be stretched at speeds of up to 0.2 nm ms⁻¹ (compare with values in Fig. 3b). Upon stretching, a protein–ligand bond will break under an applied force as described²⁶, where the rate of bond dissociation is given by $K_{off} = K_0 \exp(F \Delta x_{bond} / kT)$. K_0 is the bond off-rate in the absence of a force and Δx_{bond} is the distance over which the bond becomes unstable and fails.

To examine the effect of domain unfolding on a protein–ligand bond we simulated two mechanical arrangements: a bond with a short protein ($l_c = 50$ nm) that lacks extensible domains (Fig. 5b) and a bond with a modular protein composed of 15 FN-III domains (Fig. 5c). The bond parameters used ($\Delta x_{bond} = 0.04$ nm, and $K_0 = 0.95$ s⁻¹) were those of P-selectin⁴, an extracellular protein that supports cell rolling and that is unlikely to be extensible because of its high content of disulphide bonds⁸. The force–extension relationship for the polypeptides was calculated with the WLC model ($p = 0.42$ nm; Fig. 2a). The unfolding parameters for FN-III domains were those determined for the tenascin protein (Fig. 2).

Monte Carlo simulations of a protein–ligand bond with a protein contour length of $l_c = 50$ nm (see, for example, Fig. 5b) give force–extension curves corresponding to a simple WLC, with an abrupt ending marking bond rupture. The bond ruptured after extending the protein by 40 ± 12 nm ($n = 500$); rupture forces averaged 445 ± 283 pN ($n = 500$).

In contrast, Monte Carlo simulations of a bond with a modular protein (Fig. 5c; 15 tandem FN-III domains; folded length, 50 nm) give force–extension curves with a series of force peaks forming a saw-tooth pattern that is similar to that observed experimentally here for tenascin. In this case, the bond ruptured

after extending the protein by an average of 121 ± 94 nm; rupture forces averaged 65 ± 44 pN ($n = 500$). The lower bond-rupture forces predicted in this case suggest that, during cell rolling, protein-domain unfolding can be considered to be part of the susceptibility of the kinetics of the protein–receptor bond to an applied force (reactive compliance^{4,27}).

Received 17 November 1997; accepted 20 February 1998.

1. Palecek, S. P., Loftus, J. C., Ginsberg, M. H., Lauffenburger, D. A. & Horwitz, A. F. Integrin–ligand binding properties govern cell migration speed through cell–substratum adhesiveness. *Nature* **385**, 537–540 (1997).
2. Lauffenburger, D. A. & Horwitz, A. F. Cell migration: a physically integrated molecular process. *Cell* **84**, 359–369 (1996).
3. Alon, R., Hammer, D. A. & Springer, T. A. Lifetime of the P-selectin carbohydrate bond and its response to tensile force in hydrodynamic flow. *Nature* **374**, 539–542 (1995).
4. Alon, R. *et al.* The kinetics of L-selectin tethers and the mechanics of selectin-mediated rolling. *J. Cell Biol.* **138**, 1169–1180 (1997).
5. Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J. M. & Gaub, H. E. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science* **276**, 1109–1112 (1997).
6. Plaxco, K. W. *et al.* Rapid refolding of a proline-rich all- β -sheet fibronectin type III module. *Proc. Natl Acad. Sci. USA* **93**, 10703–10706 (1996).
7. Plaxco, K. W. *et al.* A comparison of the folding kinetics and thermodynamics of two homologous fibronectin type III domains. *J. Mol. Biol.* **270**, 763–770 (1997).
8. Bork, P. *et al.* Structure and distribution of modules in extracellular proteins. *Quart. Rev. Biophys.* **29**, 119–167 (1996).
9. Campbell, I. D. & Spitzfaden, C. Building proteins with fibronectin type III modules. *Structure* **2**, 333–337 (1994).
10. Erickson, H. P. Tenascin-C, tenascin-R and tenascin-X: a family of talented proteins in search of functions. *Curr. Opin. Cell Biol.* **5**, 869–876 (1993).
11. Chiquet-Ehrismann, R. Tenascins, a growing family of extracellular matrix proteins. *Experientia* **51**, 853–862 (1995).
12. Clark, R., Erickson, H. P. & Springer, T. A. Tenascin supports lymphocyte rolling. *J. Cell Biol.* **137**, 755–765 (1997).
13. Bork, P. & Doolittle, R. F. Proposed acquisition of an animal protein domain by bacteria. *Proc. Natl Acad. Sci. USA* **89**, 8990–8994 (1992).
14. Potts, J. R. & Campbell, I. D. Structure and function of fibronectin modules. *Matrix Biol.* **15**, 313–320 (1996).
15. Keller, T. C. S. Molecular bungees. *Nature* **387**, 233–235 (1997).
16. Tskhovrebova, L., Trinick, J., Sleep, J. A. & Simmons, R. M. Elasticity and unfolding of single molecules of the giant muscle protein titin. *Nature* **387**, 308–312 (1997).
17. Kellermayer, S. B., Smith, S. B., Granzier, H. L. & Bustanante, C. Folding-unfolding transitions in single titin molecules characterized with laser tweezers. *Science* **276**, 1112–1116 (1997).
18. Rief, M. *et al.* Single molecule force spectroscopy on polysaccharides by atomic force microscopy. *Science* **275**, 1295–1297 (1997).
19. Florin, E. L., Moy, V. T. & Gaub, H. E. Adhesion forces between individual ligand–receptor pairs. *Science* **264**, 415–417 (1994).
20. Radmacher, M., Fritz, M., Hansma, H. G. & Hansma, P. K. Direct observation of enzyme activity with the atomic force microscope. *Science* **265**, 1577–1579 (1994).
21. Aukhil, I. *et al.* Cell- and heparin-binding domains of the hexabrachion arm identified by tenascin expression proteins. *J. Biol. Chem.* **268**, 2542–2552 (1993).
22. Marko, J. F. & Siggia, E. D. Stretching DNA. *Macromolecules* **28**, 8759–8770 (1995).
23. Leahy, D. J., Hendrickson, W. A., Aukhil, I. & Erickson, H. P. Structure of a fibronectin type III domain from tenascin phased by MAD analysis of the selenomethionylprotein. *Science* **258**, 987–991 (1992).
24. Clarke, J., Hamill, S. J. & Johnson, C. M. Folding and stability of a fibronectin type III domain of human tenascin. *J. Mol. Biol.* **270**, 771–778 (1997).
25. Florin, E. L. *et al.* Sensing specific molecular interactions with the atomic force microscope. *Biosens. Bioelectr.* **10**, 895–901 (1995).
26. Bell, G. I. *et al.* Force for the specific adhesion of cells to cells. *Science* **200**, 618–627 (1978).
27. Hammer, D. A. & Apte, S. M. Simulation of cell rolling and adhesion on surfaces in shear flow: general results and analysis of selectin-mediated neutrophil adhesion. *Biophys. J.* **63**, 35–57 (1992).

Correspondence and requests for materials should be addressed to J.M.F. (e-mail: fernandez.julio@mayo.edu).

Regulation of adenovirus alternative RNA splicing by dephosphorylation of SR proteins

Arvydas Kanopka, Oliver Mühlemann, Svend Petersen-Mahrt, Camilla Estmer, Christina Öhrmalm & Göran Akusjärvi

Department of Medical Biochemistry and Microbiology, BMC, Uppsala University, Box 582, S-751 23 Uppsala, Sweden

SR proteins are a family of essential splicing factors required for early recognition of splice sites during spliceosome assembly^{1,2}. They also function as alternative RNA splicing factors when overexpressed *in vivo* or added in excess to extracts *in vitro*^{1,2}. SR proteins are highly phosphorylated *in vivo*, a modification that

is required for their function in spliceosome assembly^{3,4} and splicing catalysis^{5,6}. Here we show that SR proteins purified from late adenovirus-infected cells are inactivated as splicing enhancer or splicing repressor proteins by virus-induced dephosphorylation. We further show that the virus-encoded protein E4-ORF4 activates dephosphorylation by protein phosphatase 2A of HeLa SR proteins and converts their splicing properties into that of SR proteins purified from late adenovirus-infected cells. Taken together, our results suggest that E4-ORF4 is an important factor controlling the temporal shift in adenovirus alternative RNA splicing. We conclude that alternative pre-mRNA splicing, like many other biological processes, is regulated by reversible protein phosphorylation.

The adenovirus L1 unit⁷ is an alternatively spliced pre-mRNA in which a common 5' splice site can be joined to two alternative 3' splice sites, resulting in the production of the 52,55K (proximal 3' splice site) or the IIIa (distal 3' splice site) mRNA. During a lytic adenovirus infection, L1 alternative RNA splicing is temporally regulated, such that efficient IIIa 3' splice-site usage is confined to the late phase of virus infection⁷. We have previously shown that IIIa splicing is repressed in uninfected HeLa cell nuclear extracts (HeLa-NE) by SR proteins binding to an intronic repressor element (the 3RE)⁸ located immediately upstream of the IIIa 3' splice site.

Removing the 3RE element from a mini-52,55K–IIIa tandem *in vitro* mimics the early to late shift in L1 alternative splicing observed *in vivo* (data not shown), emphasizing that inhibition of IIIa splicing by SR proteins is important for L1 alternative 3' splice-site usage. The steady-state amount of the classical SR proteins⁹ does not change significantly during virus infection (data not shown), suggesting that the shift to increased IIIa splicing late in infection results from a virus-induced inactivation of the repressor function of SR proteins. To test this hypothesis, we compared the biological activity of SR protein from uninfected HeLa cells (SR-HeLa) and late adenovirus-infected cells (SR-Ad). Addition of excess SR-HeLa to HeLa-NE efficiently repressed IIIa splicing (Fig. 1a)⁸, whereas adding the same amount of SR-Ad to HeLa-NE resulted in only a moderate inhibition of IIIa splicing. Also, addition of SR-Ad to splicing-deficient HeLa S100 extracts¹⁰ resulted in a much smaller activation of 52,55K splicing than SR-HeLa (Fig. 1b). These results suggest that SR-Ad proteins are no longer able to function efficiently as a splicing repressor or splicing enhancer proteins, most probably as a result of losing their RNA-binding capacity (Fig. 1c). However, this reduced capacity to bind RNA does not result from a co-purification of late adenovirus RNA (data not shown), indicating that L1 splicing is not regulated by the sequestering model suggested to control E1A alternative RNA splicing¹¹.

To investigate whether the decreased activity of SR-Ad in splicing regulation and RNA binding (Fig. 1) correlated with a change in phosphorylation, *in vivo* ³²P-labelled SR-HeLa and SR-Ad proteins were analysed by two-dimensional gel electrophoresis (Fig. 2). Two differences were consistently observed in multiple experiments. First, total incorporation of ³²P into SR-Ad was significantly reduced compared with SR-HeLa. Second, with the exception of SRp20, the isoelectric point of a large fraction of the SR-Ad proteins was significantly shifted towards the basic side of the gel, suggesting that SR-Ad are underphosphorylated. That the major ³²P-labelled proteins (Fig. 2) are SR proteins was confirmed by mAb104 (ref. 12) staining of western blots (data not shown).

The adenovirus protein E4-ORF4 associates with the serine- and threonine-specific protein phosphatase 2A (PP2A)¹³. The E4-ORF4-PP2A complex has previously been suggested to regulate transcription by inducing dephosphorylation of specific transcription factors^{14,15}. To test whether E4-ORF4 also controls RNA splicing, we sought to convert the splicing properties of SR-HeLa to that of SR-Ad. Incubation of ³²P-labelled SR-HeLa in HeLa-NE in the presence of recombinant E4-ORF4 resulted in increased phosphate release compared with HeLa-NE alone (Fig. 3a). Inclusion of

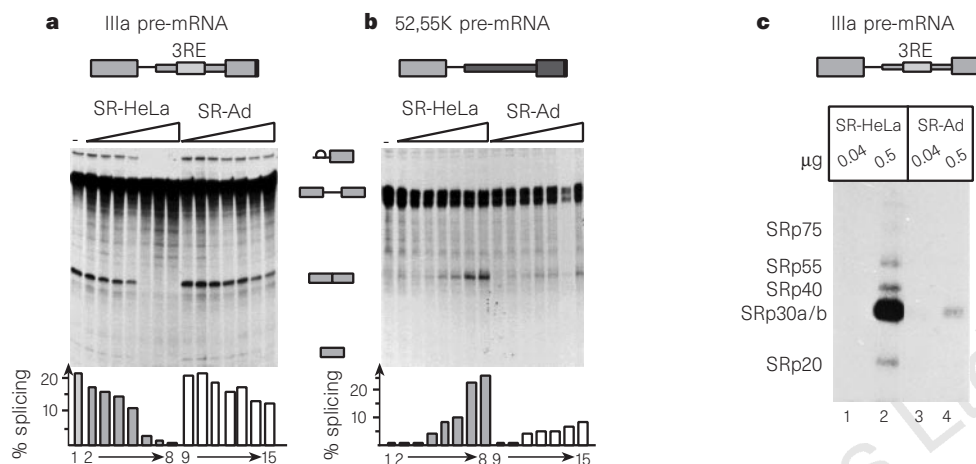


Figure 1 SR splicing factors purified from adenovirus-infected cells are functionally inactivated. Conditions for *in vitro* splicing were as described⁸. **a**, Reduced repressor activity of SR-Ad. Increasing amounts (0.1–1.0 μg) of SR-HeLa (lanes 2–8) or SR-Ad (lanes 9–15) were added to standard HeLa-NE¹⁶ splicing reactions containing the IIIa pre-mRNA⁸. Bottom, quantification of IIIa splicing efficiency. **b**, Reduced enhancer activity of SR-Ad. Increasing amounts (0.1–

1.0 μg) of SR-HeLa (lanes 2–8) or SR-Ad (lanes 9–15) were added to HeLa S100 extracts¹⁰ containing the 52,55K pre-mRNA⁸. Bottom, quantification of 52,55K splicing efficiency. **c**, Reduced RNA binding capacity of SR-Ad. Indicated amounts of SR-HeLa or SR-Ad were crosslinked to the ³²P-labelled IIIa pre-mRNA as described⁸. Products were resolved on a 12% SDS-polyacrylamide gel⁸.

okadaic acid, at a concentration that inhibits PP2A, completely annulled the effect of E4-ORF4 on SR dephosphorylation. The residual dephosphorylation activity results from phosphatases not regulated by E4-ORF4. Furthermore, preincubation of SR-HeLa with recombinant E4-ORF4 in HeLa-NE almost completely annulled the repressive effect of the SR-HeLa on IIIa pre-mRNA splicing (Fig. 3b). SR proteins inhibit IIIa splicing by blocking U2 snRNP recruitment to the spliceosome (measured as A-complex formation)⁸. E4-ORF4 alleviates the inhibitory effect of SR-HeLa on A-complex formation on the IIIa 3' splice site (Fig. 3c). Taken together, these data suggest that E4-ORF4 can convert the splicing properties of SR-HeLa proteins to that of SR-Ad. It therefore seems unlikely that an unknown factor(s) co-purifying with SR-Ad is responsible for the diminished activity of SR-Ad in splicing control (Fig. 1).

To investigate further the role of E4-ORF4 in IIIa splicing, we reproduced the effect of E4-ORF4 in a transient transfection assay. Expression of E4-ORF4 resulted in a more than 10-fold activation of IIIa pre-mRNA splicing (Fig. 4a, lanes 1 and 2), an activation that was 3RE dependent (lanes 3 and 4). Thus emphasizing that the stimulatory effect of E4-ORF4 on IIIa splicing is brought about by inactivation of SR proteins binding to the 3RE. Also co-transfection with E4-ORF4 resulted in an approximately fourfold increase in IIIa 3' splice-site usage on a mini-52,55K–IIIa tandem construct (Fig. 4b).

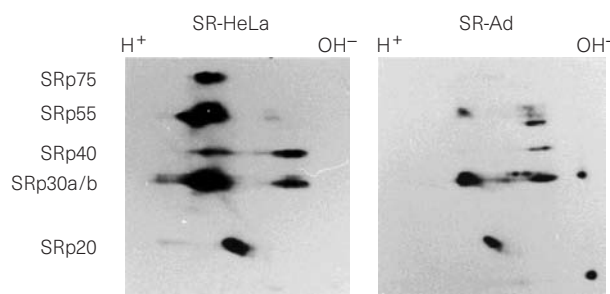


Figure 2 SR proteins purified from adenovirus late-infected cells are hypophosphorylated. Two-dimensional gel electrophoresis of SR proteins purified from uninfected or adenovirus late-infected HeLa cells metabolically labelled with ³²P-orthophosphate. H⁺ and OH⁻ denote the acidic and basic side, respectively.

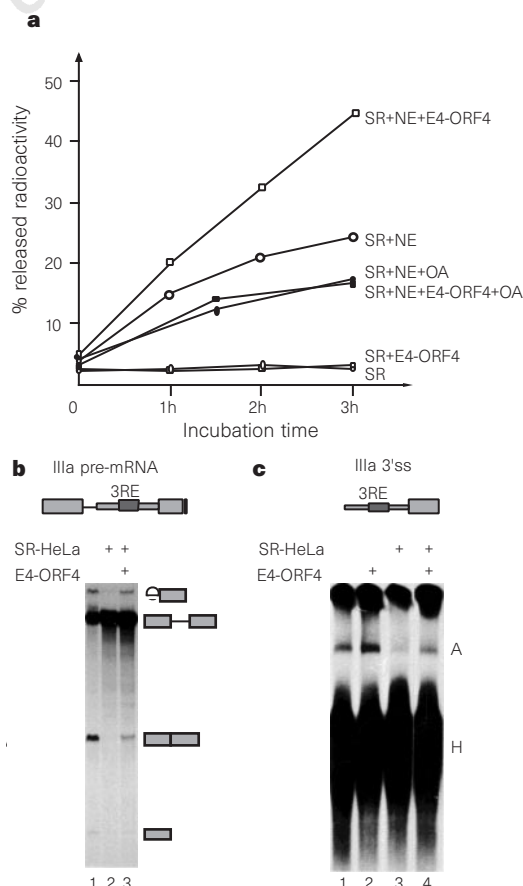


Figure 3 *In vitro* conversion of SR-HeLa to SR-Ad. **a**, E4-ORF4 induces SR protein dephosphorylation. ³²P-labelled SR-HeLa were incubated with combinations of HeLa-NE, recombinant E4-ORF4 and okadaic acid (OA). Samples were withdrawn at the indicated time points and phosphate release was measured¹⁸. **b**, E4-ORF4 rescues SR inhibition of IIIa splicing. The IIIa pre-mRNA was preincubated at 30 °C for 15 min in HeLa-NE, with or without SR-HeLa and recombinant E4-ORF4 before splicing activation (ATP addition). Conditions for *in vitro* splicing were as described⁸. **c**, Preincubation of E4-ORF4 with SR-HeLa alleviates the inhibitory effect of SR proteins on U2 snRNP recruitment to the IIIa 3' splice site (measured as A-complex formation).

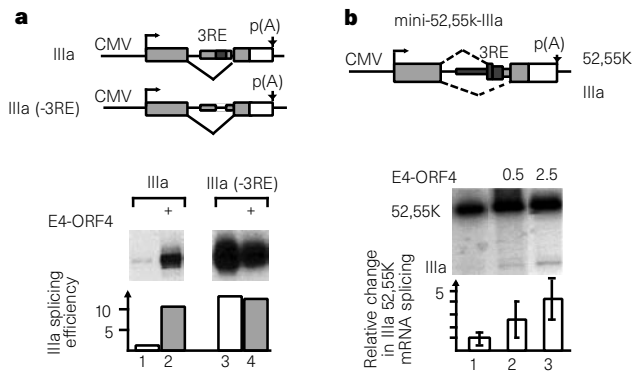


Figure 4 E4-ORF4 regulates L1 splicing in transiently transfected HeLa cells. Conditions for DNA transfection and S1 analysis were as described¹⁹. **a**, E4-ORF4 activation of IIIa splicing is dependent on 3RE. IIIa mRNA levels in HeLa cells transfected with pAdCMV-IIIa or pAdCMV-IIIa (-3RE) and empty vector (lanes 1 and 3) or pCMV-E4-ORF4 (lanes 2 and 4). Top, schematic representation of the transcription units. Bottom, quantification of IIIa splicing (results from two experiments, arbitrary units). **b**, E4-ORF4 induces a shift in L1 alternative RNA splicing. 52,55K and IIIa mRNA levels in HeLa cells transfected with pAdCMV-mini-52,55K-IIIa and empty vector (lane 1) or increasing amounts of pCMV-E4-ORF4 (lanes 2 and 3). Top, schematic representation of the transcription unit and its alternative splicing pathways. Bottom, quantification of the effect of E4-ORF4 on the ratio of IIIa/52,55K mRNA splicing (arbitrary units), with the results without E4-ORF4 set as one. Error bars indicate standard deviation, from seven experiments.

Taken together, our results suggest that the adenovirus E4-ORF4 protein functions as a virus-encoded alternative RNA splicing factor, regulating L1 IIIa 3' splice-site usage both *in vitro* and *in vivo* by modulating the activity of SR proteins. Previous results suggest that the change in L1 3' splice-site usage results from decreased 52,55K splicing combined with an increased IIIa splicing^{16,17}. Notably, E4-ORF4 activates IIIa splicing *in vivo* without having much effect on 52,55K splicing (Fig. 4b). Thus E4-ORF4 alone is unable to account for the complete shift in L1 alternative splicing⁷.

We have shown that reversible SR protein phosphorylation may be an important mechanism controlling alternative RNA splicing in mammalian cells. Thus it is possible that alternative RNA splicing might be regulated by common pathways used in signal transduction, and so it might be important in cell-cycle control, differentiation and development. □

Methods

DNA. Plasmid maps and sequences are available on the World-Wide Web at www.bmc.uu.se/IMIM/res/GA.html.

Complex assembly. Briefly, 0.5 µg of SR-HeLa were incubated with or without 4.5 pmol of His-E4-ORF4 in 15 µl HeLa-NE in buffer D for 15 min at 30 °C. Subsequently, ATP (2 mM), MgCl₂ (2.5 mM), creatine phosphate (20 mM), and substrate RNA⁸ (15 fmol) were added (final concentrations), and the reaction was incubated at 30 °C for 15 min. Heparin (2.5 µg) was added and complexes separated on a 4.25% polyacrylamide/glycerol gel¹⁶. His-E4-ORF4 was purified from *Escherichia coli* as described by the manufacturer (Novagen), except that pH 8.2 buffers were used.

Metabolic labelling of cells. Equal numbers (1.4×10^9 cells) of uninfected or adenovirus-infected HeLa cells (16 h post-infection) were resuspended (10^7 cells ml⁻¹) in phosphate-free spinner medium supplemented with 2.5% newborn calf serum. Cells were labelled with 2.5 mCi of [³²P]orthophosphate (Amersham) for 12 h. SR proteins were purified by double-salt precipitation⁹.

Two-dimensional gel electrophoresis. *In vivo* ³²P-labelled SR-HeLa or SR-Ad proteins (20 µg) were separated, on Immobiline strips with a linear pH gradient 3–10 (Pharmacia), and on a 12% SDS–polyacrylamide gel, followed by autoradiography.

SR protein dephosphorylation. Reactions (25 µl) contained 42 mM imidazole-HCl, pH 7.2, 0.5 mM EGTA, 0.25% (v/v) 2-mercaptoethanol, 2.5 mg ml⁻¹ BSA; ³²P-labelled SR-HeLa (2,000–3,000 c.p.m.) and combinations of 0.4% HeLa-NE, 0.27 µg His-E4-ORF4 and 20 nM okadaic acid. Samples were incubated at 30 °C for the indicated time and phosphate release was measured by binding to ammonium molybdate¹⁸.

Received 21 October 1997; accepted 17 February 1998.

1. Fu, X.-D. The superfamily of arginine/serine-rich splicing factors. *RNA* **1**, 663–680 (1995).
2. Manley, J. L. & Tacke, R. SR proteins and splicing control. *Genes Dev.* **10**, 1569–1579 (1996).
3. Mermoud, J. E., Cohen, P. T. W. & Lamond, A. I. Regulation of spliceosome assembly by a protein phosphorylation mechanism. *EMBO J.* **13**, 5679–5688 (1994).
4. Xiao, S.-H. & Manley, J. L. Phosphorylation of the ASF/SF2 R5 domain affects both protein-protein and protein-RNA interactions and is necessary for splicing. *Genes Dev.* **11**, 334–344 (1997).
5. Mermoud, J. E., Cohen, P. T. W. & Lamond, A. I. Ser/Thr-specific protein phosphatases are required for both catalytic steps of pre-mRNA splicing. *Nucleic Acids Res.* **20**, 5263–5269 (1992).
6. Cao, W., Jamison, S. F. & Garcia-Blanco, M. A. Both phosphorylation and dephosphorylation of ASF/SF2 are required for pre-mRNA splicing *in vitro*. *RNA* **3**, 1–12 (1997).
7. Imperiale, M., Akusjärvi, G. & Leppard, K. In *Current Topics in Microbiology* (eds Doerfler, W. & Böhm, P.) 139–171 (Verlag Chemie, Berlin, 1995).
8. Kanopka, A., Mühlemann, O. & Akusjärvi, G. Inhibition by SR proteins of splicing of a regulated adenovirus pre-mRNA. *Nature* **381**, 535–538 (1996).
9. Zahler, A. M., Lane, W. S., Stolk, J. A. & Roth, M. B. SR proteins: A conserved family of pre-mRNA splicing factors. *Genes Dev.* **6**, 837–847 (1992).
10. Krainer, A. R., Conway, G. C. & Kozak, D. The essential pre-mRNA splicing factor SF2 influences 5' splice site selection by activating proximal sites. *Cell* **62**, 35–42 (1990).
11. Himmelspach, M., Cavaloc, Y., Chebli, K., Stévenin, J. & Gattoni, R. Titration of serine/arginine (SR) splicing factors during adenoviral infection modulates E1A pre-mRNA alternative splicing. *RNA* **1**, 794–806 (1995).
12. Roth, M. B., Murphy, C. & Gall, J. G. A monoclonal antibody that recognizes a phosphorylated epitope stains lampbrush chromosome loops and small granules in the amphibian germinal vesicle. *J. Cell Biol.* **111**, 2217–2223 (1990).
13. Kleinberger, T. & Shenk, T. Adenovirus E4orf4 protein binds to protein phosphatase 2A, and the complex down regulates E1A-enhanced junB transcription. *J. Virol.* **67**, 7556–7560 (1993).
14. Bondesson, M., Ohman, K., Mannervik, M., Fan, S. & Akusjärvi, G. Adenovirus E4 open reading frame 4 protein autoregulates E4 transcription by inhibiting transactivation of the E4 promoter. *J. Virol.* **70**, 3844–3851 (1996).
15. Muller, U., Kleinberger, T. & Shenk, T. Adenovirus E4orf4 protein reduces phosphorylation of c-Fos and E1A proteins while simultaneously reducing the level of AP-1. *J. Virol.* **66**, 5867–5878 (1992).
16. Kreivi, J.-P. & Akusjärvi, G. Regulation of adenovirus alternative RNA splicing at the level of commitment complex formation. *Nucleic Acids Res.* **22**, 332–337 (1994).
17. Mühlemann, O., Kreivi, J.-P. & Akusjärvi, G. Enhanced splicing of nonconsensus 3' splice sites late during adenovirus infection. *J. Virol.* **69**, 7324–7327 (1995).
18. Ausubel, F. M. et al. in *Current Protocols in Molecular Biology* (Wiley, New York, 1995).
19. Ohman, K., Nordqvist, K. & Akusjärvi, G. Two adenovirus proteins with redundant activities in virus growth facilitates tripartite leader mRNA accumulation. *Virology* **194**, 50–58 (1993).

Acknowledgements. We thank J.-P. Kreivi for discussions and comments on the manuscript. This work was supported by the Swedish Cancer Society and the Wenner Gren Society.

Correspondence and requests for materials should be addressed to G.A. (e-mail: goran.akusjarvi@imim.uu.se).

Function of TAF_{II}-containing complex without TBP in transcription by RNA polymerase II

Elzbieta Wieczorek*†‡, Marjorie Brand*†‡, Xavier Jacq*† & László Tora*

* Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS/INSERM/ULP, BP 163, F-67404 Illkirch Cedex, CU de Strasbourg, France

‡ These authors contributed equally to this work.

Initiation of transcription of a gene from a core promoter region by RNA polymerase II requires the assembly of several initiation factors to form a preinitiation complex. Assembly of this complex^{1,2} is thought to be nucleated exclusively by the sequence-specific binding of the TFIID transcription factor complex, which is composed of the TATA-binding protein (TBP) and TBP-associated factors (TAF_{II}s) (refs 3, 4), to the different promoters.

† Present addresses: Massey Cancer Center, Virginia Commonwealth University, 401 College Street, Richmond, Virginia 23298, USA (E.W.); Columbia University, Department of Biological Sciences, 1212 Amsterdam Avenue, New York, New York 10027, USA (X.J.)

In human cells, distinct TFIID complexes exist⁵⁻⁷. They have been divided into two different subpopulations on the basis of the absence (in TFIID α) or presence (in TFIID β) of a specific TAF_{II}, hTAF_{II}30 (refs 6, 8). To characterize TFIID β complexes, we developed a two-step immunopurification procedure using antibodies raised against human (h) TAF_{II}30 and TBP (Fig. 1a). We immunopurified complexes containing hTAF_{II}30 from the flow-through fraction of a single-stranded-DNA column (Fig. 1a) with an anti-hTAF_{II}30 monoclonal antibody (1H8). Bound complexes were eluted with an excess of the corresponding epitope peptide (producing eluate 1; Fig. 1a, b). The anti-hTAF_{II}30 antibody immune-depleted all hTAF_{II}30 from the input fraction (Fig. 1b). In agreement with earlier results^{6,8}, this immunoprecipitation divided the endogenous HeLa cell TFIID complexes into two subpopulations, TFIID α and TFIID β , as the total cellular amount of two typical hTAF_{II}s, hTAF_{II}100 and hTAF_{II}55, was divided approximately in half (Fig. 1b, lanes 1, 2).

Two lines of evidence suggest that supernatant 3 contains a multiprotein complex and is not simply a product of TFIID β dissociation. First, all the proteins present in supernatant 3 must be associated directly or indirectly with hTAF_{II}30, as they have been specifically immunopurified with an anti-hTAF_{II}30 antibody; and second, not only TAF_{II}s but several other protein species, not seen normally in TFIID complexes, were also specifically present in this TFIID β -free supernatant (Fig. 1d). We called this new multiprotein complex TBP-free TAF_{II}-containing complex (TFTC).

Many TAF_{II}s, including hTAF_{II}20, hTAF_{II}30, hTAF_{II}55, hTAF_{II}80, hTAF_{II}100 and hTAF_{II}135, are apparently common to both TF_{II}C and TF_{II}ID β . In contrast, hTAF_{II}250 and hTAF_{II}28, which strongly interact with TBP, were absent from the TF_{II}C fraction (Fig. 1c). Moreover, recombinant TAF_{II}s (hTAF_{II}18, hTAF_{II}20, hTAF_{II}31,

a

NE

ssDNA Cellulose

0.15 M KCl

α -TAF_{II}30 IP

0.15 1.0 M KCl peptide elution (E₁)

SN₁

TFIID α

α -TBP IP

0.15 1.0 M KCl peptide elution (E₂)

SN₂

TFIID β

SN₃

TBP free TAF_{II} complex = TFTC

b

ssDNA Ft

SN₁(TFIID α)

E₁

1 2 3 4 5

TAF_{II}100

TAF_{II}55

TBP

TAF_{II}30

c

E₁

SN₂

SN₃(TFTC)

E₂(TFIID β)

1 2 3 4

TAF_{II}250

TAF_{II}135

TAF_{II}100

TAF_{II}80

TAF_{II}55

TBP

TAF_{II}30

TAF_{II}28

TAF_{II}20

d

E₁

SN₂

SN₃(TFTC)

Res-TF IID β

E₂(TFIID β)

1 2 3 4 5

206

106

80

50

32

25

TAF_{II}250

TAF_{II}135

TAF_{II}100

TAF_{II}80

IgG

TAF_{II}55

TBP

TAF_{II}31

TAF_{II}30

TAF_{II}28

IgGL

TAF_{II}20

e

TBP

SN₃(TFTC)

E₂(TFIID β)

1 2 3

TBP

f

rTLF

HeLa WCE

SN₃(TFTC)

106

80

32

25

TAF_{II}100

Tr-TLF

TLF

1 2 3

188 Nature © Macmillan Publishers Ltd 1998 NATURE | VOL 393 | 14 MAY 1998

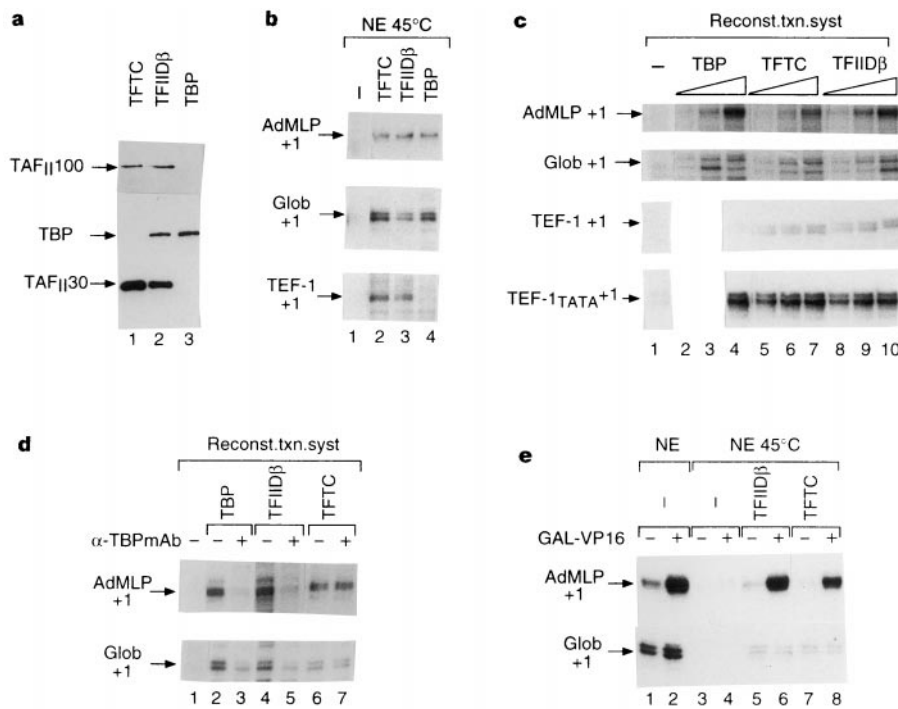


Figure 2 TFC functions in basal and activated Pol II mediated transcription. **a**, Normalization of the different fractions by immunoblotting. **b–e**, Gel photographs show products of transcription from the relevant promoters. **b**, TFC, TFIIDβ or TBP was preincubated with templates containing AdMLP, the β-globin promoter (Glob) or the TATA-less TEF-1 promoter before the addition of the heat-treated nuclear extract (NE 45°C). **c**, Increasing amounts of TBP, TFC or TFIIDβ were preincubated with the different templates before the addition of the highly purified basal factors (reconstituted transcription system, Reconst. txn. syst.). TEF-1_{TATA}, TEF-1 gene promoter with an added TATA box. **d**, Buffer or the anti-TBP monoclonal antibody (α-TBPMAb) was added to TBP, TFIID or TFC before the addition of the template DNAs and the missing factors. **e**, TFC or TFIIDβ was preincubated with the (17M)5/AdMLP and the β-globin templates in the presence or absence of the activator GAL-VP16 before addition of the heat-treated nuclear extract (NE 45°C).

TFIID or TBP in basal RNA polymerase II (Pol II) mediated transcription, such as YY1 (ref. 10) or TAF170 (refs 11, 12) (data not shown). To determine whether TFC contains a TBP-like factor(s), like the *Drosophila* TRF–nTAF complex¹³, we identified a novel human TBP-like factor (hTLF; see Methods) and generated antibodies against it (Methods). hTLF has significant homology to the core domains of both hTBP (64% similarity) and *Drosophila* TRF¹⁴ (60% similarity). Anti-hTLF antibodies recognize endogenous hTLF in HeLa cell extracts and recombinant TLF, but we found no hTLF in the immunopurified TFC (Fig. 1f). Thus, TFC does not contain any known TBP-like factor.

We then determined whether TFC can function in Pol II mediated transcription. To ensure that comparable amounts of TFC, TFIIDβ and TBP were used in the transcription reactions, we normalized these fractions by immunoblotting (Fig. 2a). TFC, TFIIDβ or TBP was first preincubated with promoter-containing DNA templates, two of which contained both a TATA box and an initiator element (the adenovirus major late promoter (AdMLP) and the rabbit β-globin gene promoters^{5,6}) and one of which was a TATA-less promoter that contained only an initiator element (the transcription-enhancer factor-1 (TEF-1) gene promoter¹⁵). A heat-inactivated nuclear extract¹⁶ was then added to the reactions for 30 min at 25°C and transcription started. Surprisingly, TFC was as efficient in supporting transcription initiation from TATA-box-containing promoters as TFIIDβ or recombinant (r) TBP (Fig. 2b). Moreover, TFC and TFIIDβ supported transcription initiation not only on TATA-containing promoters but also on the TATA-lacking promoter (Fig. 2b). In agreement with results showing that TAF_{II}s are required for transcription from TATA-less promoters¹⁷, rTBP did not allow transcription from the TATA-less TEF-1 promoter (Fig. 2b). TFC eluted from the gel-filtration column at between 600K and 1,300K also supported transcription initiation in this system (data not shown). These results indicate that TFC can function in Pol II mediated transcription. Furthermore, heat treatment of the HeLa nuclear extracts seems to inactivate the TFC activity (Fig. 2b, lane 1).

We then tested the TFC activity in a highly purified, TBP/TFIID-free, transcription system (Fig. 2c) in which the general Pol II transcription factors were either purified recombinant proteins

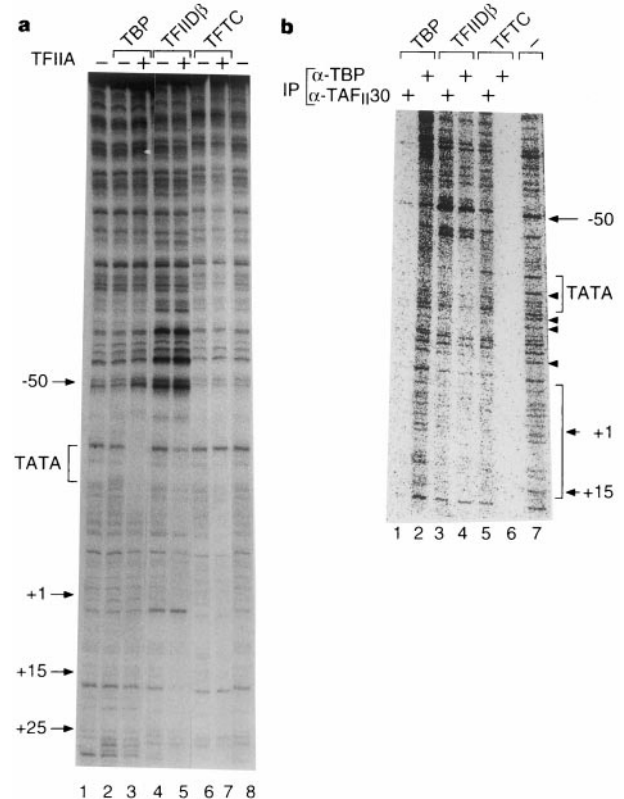


Figure 3 TFC recognizes the AdMLP. **a**, Complexes were formed on the AdMLP probe in the absence or presence of purified endogenous TFIIA²⁶. Under these conditions, TFIIA alone did not bind to the AdMLP (data not shown). **b**, TBP, TFIIDβ or TFC was incubated with either the anti-TBP (α-TBPMAb) monoclonal antibody (3G3) or the anti-hTAF_{II}30 (α-TAF_{II}30 monoclonal antibody) (2F4) and the different complexes were allowed to bind to the AdMLP probe. Complexes were immobilized, washed and briefly digested with DNase I (ref. 18). Control digestions were performed with the probe alone (–). Arrowheads and the vertical square brackets indicate regions protected by TFC and numbers are relative to the transcription start site. IP, immunoprecipitation.

(TFIIB, TFIIE α , TFIIE β and TFIIF) or extensively purified factors from HeLa cell nuclear extracts (TFIIA, TFIID and Pol II). Increasing amounts of TFTC, TFIID β and TBP were tested for their ability to support basal transcription in this system. TFTC can initiate transcription on the AdMLP and the β -globin promoter in a similar manner to TFIID β and TBP, but with three times lower efficiency (Fig. 2c). In this system, transcription initiation from the TATA-less TEF-1 gene promoter in the presence of TFTC was weak but as efficient as with TFIID β (Fig. 2c). Furthermore, when we introduced a TATA box into the TEF-1 gene promoter (TEF-1_{TATA}; ref. 15), both TFTC and TFIID β allowed more efficient transcription initiation than from the TATA-less promoter (Fig. 2c). Thus, TFTC can support transcription initiation in this TBP-free transcription system from both TATA-containing and TATA-less promoters.

To eliminate the possibility that trace amounts of TFIID or TBP contaminated the purified transcription factors and/or TFTC, we included an anti-TBP monoclonal antibody (1C2) in the transcription reactions. This antibody inhibits TBP- and TFIID-dependent Pol II transcription from TATA-containing promoters *in vitro*⁹. It also inhibited TBP- and TFIID-dependent transcription from both the AdMLP and the globin promoter in the highly purified transcription system (Fig. 2d). In contrast, it did not inhibit TFTC-dependent transcription (Fig. 2d). Thus TFTC is a new transcription factor that functions in a TBP-independent manner in transcription initiation.

Next we examined whether TFTC can mediate transcriptional activation. We first preincubated TFTC and TFIID β with the AdMLP template containing five binding sites for the transcriptional activator GAL4 (and the rabbit β -globin promoter as a control) in the presence or absence of GAL-VP16. We then added a heat-inactivated nuclear extract to the reactions for 30 min at 25 °C and transcription was started. Under these conditions, TFTC was as efficient in mediating transcriptional activation by GAL-VP16 as TFIID β (Fig. 2e).

By using DNase I footprinting, we tested where TFTC, TFIID and TBP contact the AdMLP. DNA-protein complexes were formed and cleaved by DNase I (Fig. 3a). Under the conditions used for *in vitro* transcription, there was no protection of the TATA-box region in the absence of endogenous TFIIA by TBP, TFIID β or TFTC (Fig. 3a, lanes 2, 4, 6); however, there were several hypersensitive sites seen with TFIID β alone (Fig. 3a, lane 4). Binding of TBP and TFIID β to the AdMLP was stimulated by TFIIA (Fig. 3a, lanes 3, 5). A weak extended protection, from -27 to roughly +15, was also observed when both TFTC and TFIIA were incubated with the probe (Fig. 3a; compare lanes 6, 7). To determine whether TFTC alone can bind to the promoter, we used the DNase I footprinting method with immobilized proteins¹⁸. Under these conditions, TFTC on its own bound specifically to the AdMLP (Fig. 3a, lane 5) and resulted in a protection similar to that obtained when its binding was stimulated by TFIIA. The antibodies did not immunoprecipitate protein-bound promoter fragments in reactions in which their corresponding epitopes were missing (Fig. 3b, lanes 1, 6), indicating that the binding of TFTC to the promoter was specific. In both cases, the protection and the hypersensitive-site patterns obtained with TFTC were different from those obtained with TFIID β . Thus, TFTC can bind to the AdMLP promoter and its binding is different from that of TFIID.

So far, TBP-independent Pol II transcription has only been reported with either the YY1 transcription factor¹⁰ or the *Drosophila* TBP-related factor¹³. Our results indicate that TBP-independent transcription may be more common than originally thought and may change our understanding of the regulation of eukaryotic gene expression. As TFTC recognizes sequences around the initiator element on the AdMLP, the TAF_{II}s present in TFTC may be important in promoter recognition. TAF_{II}250, TAF_{II}135, TAF_{II}100, TAF_{II}55 and TAF_{II}31 (or TAF_{II}30) are involved in the interactions between TFIID and the promoter DNA downstream of

the AdMLP TATA box¹⁹. Thus, these TAF_{II}s, but not TAF_{II}250, may have a similar role in the TFTC. The functional data obtained with TFTC further demonstrate that TAF_{II}s are vital not only in recognizing TATA-containing and TATA-less promoters but also in recruiting the other general transcription factors into the preinitiation complex^{20–22}. Moreover, basal factors seem to stabilize the binding of TFTC to the promoter. Alternatively, the unidentified protein species in TFTC may also contribute to site-specific binding of TFTC to the promoter and in formation of the preinitiation complex. As there is no TBP, TAF_{II}250 or TLF in TFTC, the binding of TFTC to the AdMLP did not induce any hypersensitive sites or strong protections around the TATA box, but resulted in a protection around the initiator region of the promoter. Our results indicate that TBP-free Pol II mediated transcription can occur in eukaryotic cells and that multiple preinitiation complexes, assembled with different TFIIDs or TFTC(s), are important in regulation of gene expression. □

Methods

Immunization and monoclonal antibody production. To generate the anti-hTAF_{II}30 antibody (1H8) that allowed the preparation of TFIID β by peptide elution, we injected mice with 100 μ g of the synthetic peptide encompassing the first 20 amino acids of hTAF_{II}30, coupled to ovalbumin. These injections were performed intraperitoneally three times at 2-week intervals. Spleen cells were fused with Sp2/0.Ag 14 myeloma cells and, at day 10 culture, we tested supernatants on endogenous hTAF_{II}30 by western blot analysis. Supernatants were also tested with HeLa cell nuclear extracts for immunoprecipitation and subsequent elution of the antigen with the epitope peptide.

The different anti-TBP monoclonal antibodies, the anti-TAF_{II}30 antibody (2F4), the anti-TAF_{II}20 antibody (22TA), the anti-TAF_{II}28 antibody (15TA), the anti-TAF_{II}55 antibody (19TA), the anti-TAF_{II}100 antibody (2D2) and the anti-TAF_{II}135 antibody (20TA) were described in refs 5, 6, 8, 9, 20, 23. The anti-TAF_{II}250 antibody was from Santa Cruz Biotechnology, and the anti-TAF_{II}80 polyclonal antibody was from Transduction Laboratories.

Identification of hTLF and antibody production. A computer scan of the available expressed-sequence-tag (EST) database revealed 11 human gene products that encode different portions of the same protein, which has significant homology to the C-terminal domains of both hTBP and dTRF. This protein has been called human TBP-like factor as it is 64% similar and 41% identical to the hTBP core and 60% similar and 38% identical to the dTRF core¹⁴ (sequence alignment can be obtained upon request). One of the EST complementary DNA clones (accession number AA412574) was obtained from the IMAGE Consortium²⁴ and inserted in the *Hind*III–*Not*I sites of the pET32b vector, in frame with the thioredoxin protein, and the recombinant fusion protein (rTr- Δ TLF) was overexpressed in *Escherichia coli*. Next the synthetic peptide EIRLPETKNNRPHYASYEPELH, from amino acids 114–135 of hTLF, was coupled to ovalbumin. Mice were injected intraperitoneally, three times at 2-week intervals, with 100 μ g synthetic peptide. The polyclonal mouse sera were tested on endogenous hTLF, on hTLF-transfected COS cell extracts and on *E. coli*-expressed rTr- Δ TLF by western blot analysis.

Immunoprecipitation. Monoclonal antibodies for immunoprecipitation were purified for ascites fluids using caprylic acid precipitation followed by precipitation with 50% ammonium sulphate, and were dialysed. Immunoprecipitation was performed as described⁵. HeLa cell nuclear extract was first passed through a single-stranded-DNA cellulose column to eliminate proteins that bind non-specifically to the resin used later in immunoprecipitations. Fractions were immunoprecipitated with monoclonal antibodies bound to protein G-Sepharose (Pharmacia), as indicated in the figures. The complexes bound to protein G-Sepharose/antibody were washed three times with IP buffer (25 mM Tris-HCl, pH 7.9, 10% (v/v) glycerol, 0.1% NP40, 0.5 mM dithiothreitol and 5 mM MgCl₂) containing 0.5 M KCl and two times with IP buffer containing 100 mM KCl. After washing, bound proteins were eluted with a 200–1,000-fold excess of the corresponding epitope peptide and analysed by SDS-PAGE. The gels were either silver-stained or transferred to nitrocellulose membrane and probed with the indicated antibodies.

In vitro transcription. *In vitro* transcription reactions were assembled using heat-treated HeLa cell nuclear extract (45 °C for 15 min) or the highly purified

general Pol II transcription factors^{7,20,25,26}. All protein fractions were dialysed against buffer B (25 mM Tris-HCl, pH 7.9, 50 mM KCl, 0.5 mM dithiothreitol, 0.1 mM EDTA and 20% glycerol (v/v)). 25- μ l reactions contained either 25 ng 17M/5pAL7 and pG1 (ref. 5) or 100 ng pTEF(Δ -138) or pTEF(Δ -138_{TATA})¹⁵, with aliquots of TFIIA, TFIID β and recombinant TBP⁵. Where indicated, 200 ng purified anti-TBP monoclonal antibody 1C2 was also included in the reactions before the other factors were added. GAL-VP16-activated transcription was performed as described⁵. After the preincubation steps (30 min), transcription was initiated by addition of nucleoside triphosphates to 0.5 mM and MgCl₂ to 5 mM. Transcriptions were incubated at 25°C for 45 min. Correctly initiated transcripts from the different promoters were analysed by quantitative S1 nuclease analysis^{15,27}.

DNase I footprinting. DNase I footprinting was performed as described^{18,19}. The labelled AdMLP-containing probes were amplified by polymerase chain reaction on either the 17M5/pAL7 (ref. 28) (Fig. 3a) or the pM677 (ref. 29) (Fig. 3b) templates. For the footprinting experiments, ten times more TBP, TFIID β , and TFIIA was used than in the transcription reactions.

Received 4 December 1997; accepted 6 March 1998.

1. Zavel, L. & Reinberg, D. Common themes in assembly and function of eukaryotic transcription complexes. *Annu. Rev. Biochem.* **64**, 533–561 (1995).
2. Roeder, R. G. The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem. Sci.* **21**, 327–335 (1996).
3. Verrijzer, C. P. & Tjian, R. TAFs mediate transcriptional activation and promoter selectivity [see comments]. *Trends Biochem. Sci.* **21**, 338–342 (1996).
4. Chang, M. & Jaehning, J. A. A multiplicity of mediators: alternative forms of transcription complexes communicate with transcriptional regulators. *Nucleic Acids Res.* **25**, 4861–4865 (1997).
5. Brou, C. *et al.* Distinct TFIID complexes mediate the effect of different transcriptional activators. *EMBO J.* **12**, 489–499 (1993).
6. Jacq, X. *et al.* Human TAFII30 is present in a distinct TFIID complex and is required for transcriptional activation by the estrogen receptor. *Cell* **79**, 107–117 (1994).
7. Bertolotti, A., Lutz, Y., Heard, D. J., Chambon, P. & Tora, L. hTAFII68 a novel RNA/SSDNA-binding protein with homology to the pro-oncoproteins TLS/FUS and EWS is associated with both TFIID and RNA polymerase II. *EMBO J.* **15**, 5022–5031 (1996).
8. Mengus, G. *et al.* Cloning and characterization of hTAFII118, hTAFII20 and hTAFII28; three subunits of the human transcription factor TFIID. *EMBO J.* **14**, 1520–1531 (1995).
9. Lescure, A. *et al.* The N-terminal domain of human TATA-binding protein plays a role in transcription from TATA-containing RNA polymerase II and III promoters. *EMBO J.* **13**, 1166–1175 (1994).
10. Ushuva, A. & Sherk, T. TATA-binding protein-independent initiation: YY1, TFIIB, and RNA polymerase II direct basal transcription on supercoiled template DNA. *Cell* **76**, 1115–1121 (1994).
11. Timmers, H. T. M., Meyers, R. E. & Sharp, P. A. Composition of transcription factor B-TFIID. *Proc. Natl Acad. Sci. USA* **89**, 8140–8144 (1992).
12. Van der Knaap, J. A., Willem Borst, J., van der Vliet, P. C., Gentz, R. & Timmers, H. T. M. Cloning of the cDNA for the TATA-binding protein-associated factor₁₇₀ subunit of transcription factor B-TFIID reveals homology to global transcription regulators in yeast and *Drosophila*. *Proc. Natl Acad. Sci. USA* **94**, 11827–11832 (1997).
13. Hansen, S. K., Takada, S., Jacobson, R. H., Lis, J. T. & Tjian, R. Transcription properties of a cell type specific TATA-binding protein, TRF. *Cell* **91**, 71–83 (1997).
14. Crowley, T. E., Hoey, T., Liu, J. K., Jan, Y. N., Jan, L. Y. & Tjian, R. A new factor related to TATA-binding protein has highly restricted expression patterns in *Drosophila*. *Nature* **361**, 557–561 (1993).
15. Boam, D. S., Davidson, I. & Chambon, P. A TATA-less promoter containing binding sites for ubiquitous transcription factors mediates cell type-specific regulation of the gene for transcription enhancer factor-1 (TEF-1). *J. Biol. Chem.* **270**, 487–494 (1995).
16. Nakajima, N., Horikoshi, M. & Roeder, R. G. Factors involved in specific transcription by mammalian RNA polymerase II: purification, genetic specificity, and TATA box-promoter interactions of TFIID. *Mol. Cell. Biol.* **8**, 4028–4040 (1988).
17. Pugh, B. F. & Tjian, R. Transcription from a TATA-less promoter requires a multisubunit TFIID complex. *Genes Dev.* **5**, 1935–1945 (1991).
18. Purnell, B. A., Emanuel, P. A. & Gilmour, D. S. TFIID sequence recognition of the initiator and sequences further downstream in *Drosophila* class II genes. *Genes Dev.* **8**, 830–842 (1994).
19. Oelgeschlaeger, T., Chiang, C. M. & Roeder, R. G. Topology and reorganization of a human TFIID-promoter complex. *Nature* **382**, 735–738 (1996).
20. Dubrovskaya, V. *et al.* Distinct domains of hTAFII100 are required for functional interaction with transcription factor TFIIFb (RAP30) and incorporation into the TFIID complex. *EMBO J.* **15**, 3702–3712 (1996).
21. Ruppert, S. & Tjian, R. Human TAFII250 interacts with RAP74: implications for RNA polymerase II initiation. *Genes Dev.* **9**, 2747–2755 (1995).
22. Hisatake, K. *et al.* Evolutionary conservation of human TATA-binding-polypeptide-associated factors TAFII31 and TAFII80 and interactions of TAFII80 with other TAFs and with general transcription factors. *Proc. Natl Acad. Sci. USA* **92**, 85–89 (1995).
23. Lavigne, A. C. *et al.* Multiple interactions between hTAFII55 and other TFIID subunits. Requirements for the formation of stable ternary complexes between hTAFII55 and the TATA-binding protein. *J. Biol. Chem.* **271**, 774–780 (1996).

24. Lennon, G., Auffray, C., Polymeropoulos, M. & Soares, M. B. The I.M.A.G.E. Consortium: an integrated molecular analysis of genomes and their expression. *Genomics* **33**, 151–152 (1996).
25. Gerard, M. *et al.* Purification and interaction properties of the human RNA polymerase B(II) general transcription factor BTF2. *J. Biol. Chem.* **266**, 20940–20945 (1991).
26. De Jong, J. & Roeder, R. G. A single cDNA, hTFIIA/alpha, encodes both the p35 and p19 subunits of human TFIIA. *Genes Dev.* **7**, 2220–2234 (1993).
27. Tora, L. *et al.* The human estrogen receptor has two independent nonacidic transcriptional activation functions. *Cell* **59**, 477–487 (1989).
28. Brou, C. *et al.* Different TBP-associated factors are required for mediating the stimulation of transcription *in vitro* by the acidic transactivator GAL-VP16 and the two nonacidic activation functions of the estrogen receptor. *Nucleic Acids Res.* **21**, 5–12 (1993).
29. Moncollin, V., Miyamoto, N. G., Zheng, X. M. & Egly, J. M. Purification of a factor specific for the upstream element of the adenovirus-2 major late promoter. *EMBO J.* **5**, 2577–2584 (1986).

Acknowledgements. We thank P. Chambon for support; J. C. Dantonel for help in identification and cloning of hTLF; E. Scheer for technical assistance; Y. Lutz for antibodies; D. Boam, V. Dubrovskaya, A. C. Lavigne, G. Mengus, I. Davidson and the IMAGE Consortium for reagents; H. T. M. Timmers for antibodies and for discussing unpublished results; A. Bertolotti for discussions; D. J. Heard for discussions and reading the manuscript; P. Eberling for peptide synthesis; the cell culture group for HeLa cells; R. Buchert, J.-M. Lafontaine and B. Boulay for illustrations; and A. Ozyhar for his contribution to the training of E.W. E.W. was supported by a fellowship from the Ministère de l'Enseignement Supérieur et de la Recherche. Research was supported by grants from the CNRS, the INSERM, the Hôpital Universitaire de Strasbourg, the Ministère de la Recherche et Technologie, the Fondation pour la Recherche Médicale and the Association pour la Recherche contre le Cancer.

Correspondence and requests for materials should be addressed to L.T. (e-mail: laszlo@titus.u-strasbg.fr).

corrections

Structure of the $\alpha\beta$ tubulin dimer by electron crystallography

Eva Nogales, Sharon G. Wolf & Kenneth H. Downing

Nature **391**, 199–203 (1998)

In this Letter, the numbers for the secondary structure elements involved in Taxol binding are incorrect (page 202, second-to-last paragraph of main text). The sentences giving the correct numbers are, “In our model, the C-3’ is near the top of helix H1 (that is, between β :15–25), and the C2 group near H6 and the H6–H7 loop (that is, between β :212–222). The main interaction of the taxane ring is at L275, at the beginning of the B7–H9 loop.” □

Spatial and temporal organization during cardiac fibrillation

Richard A. Gray, Arkady M. Pertsov & José Jalife

Nature **392**, 75–78 (1998)

The x-axis of Fig. 1d was mislabelled: the frequency values should instead read 0, 10, 20, 30, 40 Hz. □

Managing microbes

This listing of new product announcements covers microbiology, and includes 1 tools, a CO₂ incubator, an autoclave, information on microfiltration for epifluorescence

Biology system

From Don Whitley Scientific

A system for bacterial and yeast identification

This system is available in automated or manual formats with the capability to identify over 1,100 species and groups, and can be used for microbiology studies in clinical, veterinary, plant and environmental sciences. All testing follows the same simple protocol, standardizing laboratory procedures and eliminating the need to use a range of tests for different organisms, says Don Whitley. A suspension of microbes is simply added to the microplates, which are supplied prepared. Each well contains one of 95 different carbon-based substrates. Moreover, there are MicroPlate test panels for Gram-negative, Gram-positive and lactic-acid bacteria and yeasts. The method for preparing all plates is identical, requiring only inoculation with a standardized suspension of organisms.

Reader Enquiry No. 100

Cardinal

From Optomax

A bacterial colony-counting system that is designed to eliminate manual adjustments

This high-throughput, automatic colony counting system has been upgraded with software that controls the video camera lens (to eliminate manual adjustments), counts Petri film and interfaces directly with Microsoft Excel software. It features a plate viewer, a high-resolution video camera and Windows-based software to control lens operation and bright- or dark-field illumination. Designed to analyse bacterial or mammalian cultures in a wide variety of growth media, this system can count Petri film and perform analysis of up to 600 plates per hour. Results are provided as counts, counts per plate, counts per millilitre, area and frame area. The system will also provide an audit trail.

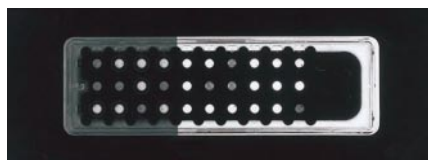
Reader Enquiry No. 101

BBL Crystal

From Becton Dickinson

A complete microbial identification system

Like all other members of this family of products, these three new panels for Gram-positive, rapid Gram-positive and *Neisseria/Haemophilus* testing offer total laboratory standardization, says the company. This includes the convenience of one-step automatic inoculation, no reagent additions or oil overlays, a computerized database, and enhanced safety and accuracy. The Gram-



Becton Dickinson's BBL Crystal additions.

positive identification system is said to cover 121 organisms, providing four-hour identification of the more common isolates. There is also a four-hour identification kit for 35 *Neisseria/Haemophilus* species. BBL Crystal consolidates identification techniques, using both chromogenic and fluorogenic substrates to reduce laboratory costs, labour and total test time. There are also enteric/nonfermenter, rapid-stool enteric and anaerobe systems. All of the systems feature the same protocol — the user simply pours the inoculum directly into the device, rocks it to fill the test wells and then snaps on the lid. No pipettes or reagent additions are necessary, and the device stays closed and secure. Convenient, stackable ten-device trays save incubator space, and the colour-change and fluorescence reactions are said to be easy to read and process.

Reader Enquiry No. 102

Atmosphere generation systems

From Oxoid

Three new additions to the atmosphere generation systems range

For simple gas generation, Oxoid has introduced the CO₂Gen, CO₂Gen Compact and CampyGen Compact systems for producing the appropriate incubation conditions for microorganisms. The final atmosphere for the two CO₂ generation systems is approximately 5 per cent CO₂ by volume, which is said to be suitable for growth of many fastidious species, such as *Haemophilus* and *Neisseria*. The O₂ level is somewhat lower at 16 per cent with these systems. The CampyGen Compact is designed so that a small number of plates can be cultured under microaerophilic conditions. The final atmosphere is reduced to 6 per cent O₂ and is supplemented with approximately 14 per cent CO₂ for growth of *Campylobacter* species, and other microaerophilic organisms. Both Compact products utilize a sachet contained within a bag-and-clip system, and 1–2 plates can be cultured in each transparent bag. A see-through compact bag system allows plates to be observed for growth after 24 h, without affecting the culture of more slowly growing organisms.

Reader Enquiry No. 103

Fluorescence and microscopy

Microbiology reference

From Corning Separations Division

A new reference for staining microorganisms using epifluorescence microscopy

Created as a reference tool for those using membrane-based microfiltration with epifluorescence microscopy, this is a resource to assist researchers with stain and membrane selection. Entitled *Staining Microorganisms for Epifluorescent Detection and Enumeration*, this publication begins with a discussion of the basics, followed by more detailed information on general, organism-specific and physiological stains. A step-by-step EpiCount Method for the epifluorescence application is included.

Reader Enquiry No. 104

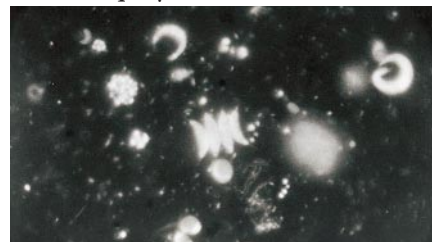
Epifluorescence microscopy

From Osmonics

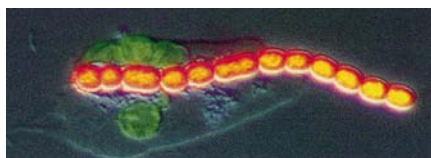
Direct observation and counting of viable and non-viable microorganisms in >30 min

According to the manufacturer, this technique compares favourably with traditional culturing methods, which may require incubation times of up to 72 hours. Traditional methods can underestimate the total number of microorganisms due to the selective nature of the media employed, and the fact that they do not account for non-viable microorganisms or the tendency of microorganisms to form clumps or microcolonies. The epifluorescent microscopy technique is said to represent the total population of viable and non-viable microorganisms, providing improved sensitivity over traditional culturing and pour-plate procedures. Poretics black polycarbonate membrane filters are used to provide the same advantages as regular polycarbonate membranes, but have been specially treated to give a very low autofluorescence, says Osmonics. This characteristic permits the observation of high viability of microorganisms on the surface of the membrane against a non-distracting background.

Reader Enquiry No. 105



Low-background epifluorescence from Osmonics.



Syto 13 live-cell nucleic acid stain shows phagocytosis of *Anabaena* sp. by a neutrophil.

Fluorescent probes

From Molecular Probes

Reagents, kits and fluorescence techniques for bacteriology, cell biology and microbiology

Molecular Probes offers nearly 3,000 fluorescent probes and other reagents for work with microorganisms or animal cells. Many are sensitive at the single-organism level and are easily adapted to high-throughput, automated analysis. The Syto stain kits use visible light-excitable stains and penetrate most bacterial species to label DNA and RNA with bright green or red fluorescence and low background. There is a bacterial counting kit for flow cytometric enumeration of bacteria. The SytoX green nucleic acid stain penetrates only compromised bacterial and animal cell membranes and exhibits >500-fold fluorescence enhancement upon binding to nucleic acids, states the manufacturer. There are also bacterial viability kits compatible with microscopy, fluorometry and flow cytometry, as well as kits for distinguishing Gram-positive and Gram-negative bacteria and showing whether cells are intact or permeabilized. For detecting contamination of cell cultures, there is a kit with three fluorescent dyes to detect and identify fungal and bacterial contaminants. There is also the MycoFluor kit to detect *Mycoplasma* contamination in cell cultures.

Reader Enquiry No. 106

Cobra

From Vision Engineering

A microscope system that uses expanded-pupil (EP) technology

Designed for less fatigue in microscopy, the microscope uses enlarged eyepieces and increases the distance between the operator's



The Cobra's hood — an answer to fatigue.

eye and the eyepiece, allowing more head movement in all three planes. Coupled with high-quality optics, EP viewing enables the user to comfortably view high-resolution images with less eye, neck and back fatigue, says Vision. The system features zoom magnification up to $\times 160$ with a working distance to 312 mm. Refractive-disk technology and a straight optical path improves resolution at high magnification, as well as the depth of field and colour rendition. A wide range of mounting, lighting and photomicrography options is available.

Reader Enquiry No. 107

Around the bench

Century autoclave

From Steris

For sterilizing a wide range of laboratory products and infectious waste

The system offers a touch-sensitive screen control, an 18-cycle programmer, integral cycle print-out and non-lubricated door seals for ease-of-use, reliability and enhanced operator safety. It features a 'hands-free' powered vertical sliding door that is operated by a foot pedal, freeing both hands for loading and unloading. The unit's door seal has been designed so that no messy silicon or grease lubricant is necessary, and is guaranteed to be replacement-free for two years. The 316 stainless-steel, fully jacketed chamber is also guaranteed and, according to Steris, users can expect at least 15 years of trouble-free performance. Its modular piping system has fewer components and connectors, which is said to reduce maintenance and service time. It is available in sizes from 50–250 l and with gravity, pre-vacuum and isothermal options, and 30-min cycle time, using a fast cooling facility.

Reader Enquiry No. 108

Flexible shakers

From Gerhardt

Designed with a vertical stacking system to increase load capacity, without expanding the original footprint of the machine

According to the manufacturer, these CE-compliant units can routinely handle loads of up to 30 kg with either reciprocal or orbital motions available. Models are available in two sizes: the small 2 series units



No walking with Gerhardt shakers.

have a 20-mm throw and the large 5 series units have a 50-mm throw. All the shakers are programmable to the nearest r.p.m. and minute, and at the end of the run the shaker will automatically stop, allowing unattended operation. A wide range of attachments is available to enable many types of container to be used, as well as custom-made options to accommodate almost any need. An incubated shaker completes the range, and uses the same compact, efficient silent-motor system as the other shakers. This system is said to provide a low centre of gravity, reducing 'walk' during use.

Reader Enquiry No. 109

Water baths

From Techne

A range of twenty water baths that operate over the range from -40°C to $+250^{\circ}\text{C}$ with a capacity from 8–48 litres

Based on the company's unheated baths and Tempette and Tempunit thermoregulators, these units can be used either as open water baths or as closed circulating systems for external apparatus. The thermoregulators bridge to one of four sizes of water bath. Moreover, these baths incorporate carrying handles, seam-free stainless steel tanks for easy cleaning and stoved enamel steel outer cases. The four immersion thermoregulators offer a choice of temperature ranges, temperature stability and analog or digital controls. A circulating pump with external corrections, adjustable over-temperature and low-liquid-level cut-outs are standard on all models. A cooling coil is an optional extra on the two lower models, but standard on the two higher models. At the bottom end of the range, the Tempette offers a temperature range of -20°C to 95°C and covers most routine applications. The top model operates at temperatures from -40°C to 250°C , which are selected and displayed digitally. There is an audible alarm, a dual-pressure and suction pump and a RS232 port.

Reader Enquiry No. 110

ADVERTISEMENT

NEWCASTLE UNIVERSITY
MOLECULAR BIOLOGY UNIT



PROTEIN SEQUENCING

Membrane-bound, solid or solution samples

£12 per cycle

5 cycles minimum - no set up charge

For further information contact Dr. Joe Gray
Phone: +44 191 222 8612
Fax: +44 191 222 8129
Email: JOE.GRAY@newcastle.ac.uk

READER ENQUIRY NO. 14

Galaxy S

From R S Biotech

A CO₂ incubator for cell biology and tissue culture applications

Using a front-door keypad to improve operating efficiency, this incubator precisely controls temperature, CO₂ content and humidity within the specifications of current regulations. There are also alarm and temperature cut-out facilities included as standard. A dry-wall heat profiling is said to eliminate the need for a fan, which should benefit cell-biology and tissue-culture applications. Air is moved by thermodynamic convection and the response to change is said to be very rapid, for stable temperature with minimal gradient. Convection is said to dramatically reduce the potential for contamination and improve culture conditions by maximizing CO₂ dispersion, reducing sample evaporation and eliminating much of the vibration that can impede fragile cell adhesion. Solid-state, infrared CO₂ sensing helps to ensure precision control of CO₂ concentrations, and is unaffected by humidity (maintained at a stable 95 per cent relative humidity). The unit offers a 170-l internal volume, and there is an

option for three inner glass doors, which can be accessed separately and help reduce temperature variation that occurs when opening the main door.

Reader Enquiry No. 111

Microcide SQ

From Global Biotechnologies

A dual quaternary ammonium disinfectant and cleaner concentrate

This formulation has been (US) EPA registered to kill 108 different bacteria, yeast, fungi and viruses. According to the manufacturer, it has been shown to kill 24 different viruses, including cytomegalovirus, HIV-1 (AIDS) and HBV (hepatitis B), as well as T1 and T4 bacteriophage. Much of the testing was done on clinical isolates from ongoing infections and in the presence of five per cent serum and 300 p.p.m. hard water to help comply with OSHA blood-borne pathogen regulations. Microcide SQ can effectively clean and disinfect work areas, equipment, instruments, as well as hoods and incubators. The product can also be used as an ultrasonic cleaning and disinfectant solution and to presoak instruments. When diluted, it

has a pH of 8.9 and is safe for use on fabric, rubber, glass, plastic and metals, including stainless steel and aluminium.

Reader Enquiry No. 112

Internet accessible

Sheffield website

From Quest International

Website on high-grade media components, such as peptones and yeast extracts

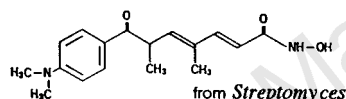
This Internet site provides easy access to up-to-date product specifications and information on almost all fermentation and cell tissue culture products in the Quest product range: www.sheffield-products.com. A search engine is incorporated to enable users to select a product based on specified data. Such information includes physical and chemical information, an amino-acid profile or a microorganism product that must be suitable for *Escherichia coli*, for example.

Reader Enquiry No. 113

These notes are compiled by Brendan Horton from information provided by the manufacturers. For more details, fill in the reader service card bound inside the journal.

ADVERTISEMENTS

Trichostatin A



Catalog No. 204-11991

This histone deacetylase inhibitor is highly effective at low concentrations, with high specificity. For more information, please call
Wako BioProducts: (800) 992-9256; (804) 271-7677.

READER ENQUIRY NO. 98

BACs

YACs

PACs

cDNAs

Research Genetics provides clones and colony membranes for a number of human, mouse, and rat libraries as well as a custom screening service for the libraries.

Research Genetics, Inc.
U.S. or Canada 800-533-4363
U.K. 0-800-89-1393
FAX 256-536-9016

READER ENQUIRY NO. 30

ANTIBODIES

For Research & Routine Histology

Inflammation – Angiogenesis – Tumors
 Leukocytes – Endothelial Cells
 Pathological Markers

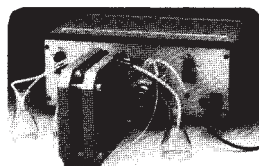
Custom Monoclonals and Polyclonals

BMA Biomedicals Ltd.
 (800) 634-3183 (US)
 Intl: +41 61 811 62 22
 +41 61 811 60 06 (fax)
 info@bma.ch

READER ENQUIRY NO. 287

LIPOSOMES

REPRODUCIBLE – HOMOGENEOUS – EXTREMELY RAPID



Our new LIPOSOMAT guarantees rapid (5ml within 60 minutes) preparation of uniformly sized unilamellar liposomes. Liposome size is selectable between 25 and ca. 600 nm diameter. Sample volume is normally 5–10ml or up to 200ml if used in combination with an additional instrument. Lipid concentration can be up to 300 mg/ml.

DIANDORM

P.O. Box 650126, D-81215 Munich
 For USA call 1-800 Liposome

READER ENQUIRY NO. 327



Join the Institute of Biology and enjoy

- ◆ journals & books ◆
- ◆ careers information ◆
- ◆ meetings & conferences ◆
- ◆ networking opportunities ◆
- ◆ shaping bioscience policies ◆
- ◆ European Biologist status *EurBiol* ◆
- ◆ Chartered Biologist status *CBiol* ◆

mention *Nature* for a
10% DISCOUNT

20 Queensberry Place
 London SW7 2DZ

+44 (0)171 581 8333

Fax: +44 (0)171 823 9409

e-mail: info@iob.primex.couk
<http://www.iob.org>

READER ENQUIRY NO. 489